POPULATION ANALYSIS WITH MICROCOMPUTERS

Volume I

PRESENTATION OF TECHNIQUES

by

EDUARDO E. ARRIAGA

with associates:
PETER D. JOHNSON, Software Associate
ELLEN JAMISON, Editorial Associate

November 1994
A comprehensive civil registration system has a number of advantages over the stop gap measures available for obtaining vital statistics. It provides a continuous flow of data. Data required for planning, administration, and research are provided for the country as a whole and its geographic subdivisions. Data are free of sampling errors. However most developing countries do not have an adequate civil registration system which can provide the vital statistics needed, and it is likely that it will be many years before all countries achieve the levels of completeness and accuracy now enjoyed by the industrialized countries.

The need for continued efforts to improve civil registration systems should not be ignored. However, until levels of completeness and accuracy do improve in the developing world, manuals such as this serve the highly commendable purpose of encouraging the analysis of available population information. The International Institute for Vital Registration and Statistics is pleased to have been able to facilitate the publication of these two valuable volumes.
# TABLE OF CONTENTS

## VOLUME I

### PREFACE

xi

### ACKNOWLEDGEMENTS

xiii

### Chapter I

1

### INTRODUCTION

1

#### Volume I

2

- Topics Covered

2

#### Volume II

3

- How to Use This Manual

4

- Some Guidelines for Using the Spreadsheets

5

### Appendix I-1

7

### Population Analysis Spreadsheets

7

### Chapter II

11

### AGE AND SEX COMPOSITION

11

#### This Chapter in Brief

11

- Spreadsheets and Methods That Can Be Used for Analyzing the Population Age Structure, According to the Available Information

12

- Where to Find the Software

13

- Introduction

13

- Age Structure as a View of the Past

13

- Checking Census Data for Consistency

14

- Detecting age misreporting

16

- Digit preference

18

- Age ratios

19

- Sex ratios

20

- Age-sex accuracy index

21

- Correcting for Age Misreporting

23

- Smoothing without modifying the totals

23

- Smoothing modifying the totals

25

- Selection of a smoothing technique

25

- Strong smoothing

26

- Other age structure considerations

33

- Open-ended age group

33

- Population adjustment

33

- Population interpolation

34

- Population shifting

34

### Appendix II-1

35

### Graphing Population Age Structures

35

### Age Pyramid

35
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population by Year of Birth</td>
<td>35</td>
</tr>
<tr>
<td>Software</td>
<td>35</td>
</tr>
<tr>
<td>Appendix II-2</td>
<td>37</td>
</tr>
<tr>
<td>Analyzing and Smoothing the Age Distribution of a Population</td>
<td>37</td>
</tr>
<tr>
<td>Analyzing the Age Structure</td>
<td>37</td>
</tr>
<tr>
<td>Age ratios</td>
<td>38</td>
</tr>
<tr>
<td>Sex ratios</td>
<td>38</td>
</tr>
<tr>
<td>Smoothing the Age Distribution</td>
<td>38</td>
</tr>
<tr>
<td>Smoothing without modifying the totals of each 10-year age group</td>
<td>39</td>
</tr>
<tr>
<td>Carrier-Farrag formula</td>
<td>39</td>
</tr>
<tr>
<td>Karup-King-Newton formula</td>
<td>39</td>
</tr>
<tr>
<td>Arriaga formula</td>
<td>40</td>
</tr>
<tr>
<td>Smoothing Modifying the Totals of Each 10-Year Age Group</td>
<td>41</td>
</tr>
<tr>
<td>United Nations formula</td>
<td>41</td>
</tr>
<tr>
<td>Strong smoothing formula</td>
<td>42</td>
</tr>
<tr>
<td>Software</td>
<td>42</td>
</tr>
<tr>
<td>Appendix II-3</td>
<td>43</td>
</tr>
<tr>
<td>Contingency Table</td>
<td>43</td>
</tr>
<tr>
<td>Appendix II-4</td>
<td>45</td>
</tr>
<tr>
<td>Estimating the Age Structure of the Open-Ended Age Group</td>
<td>45</td>
</tr>
<tr>
<td>Appendix II-5</td>
<td>49</td>
</tr>
<tr>
<td>Adjusting the Age Structure of the Population to a New Total Population</td>
<td>49</td>
</tr>
<tr>
<td>Appendix II-6</td>
<td>51</td>
</tr>
<tr>
<td>Interpolation Between Two Population Age Structures</td>
<td>51</td>
</tr>
<tr>
<td>Appendix II-7</td>
<td>53</td>
</tr>
<tr>
<td>Moving the Population to a Desired Date</td>
<td>53</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>55</td>
</tr>
<tr>
<td>Chapter III</td>
<td>57</td>
</tr>
<tr>
<td>MORTALITY</td>
<td>57</td>
</tr>
<tr>
<td>This Chapter in Brief</td>
<td>57</td>
</tr>
<tr>
<td>Direct estimation</td>
<td>57</td>
</tr>
<tr>
<td>Indirect estimation</td>
<td>57</td>
</tr>
<tr>
<td>Smoothing mortality rates</td>
<td>58</td>
</tr>
<tr>
<td>Spreadsheets, United Nations Programs, and Methods That Can Be Used</td>
<td>58</td>
</tr>
<tr>
<td>for Analyzing Mortality, According to the Available Information</td>
<td>58</td>
</tr>
<tr>
<td>Topic</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Direct Standardization</td>
<td>119</td>
</tr>
<tr>
<td>Differences Between Two Rates</td>
<td>123</td>
</tr>
<tr>
<td>Indirect Standardization</td>
<td>125</td>
</tr>
<tr>
<td>Appendix III-2</td>
<td>129</td>
</tr>
<tr>
<td>Construction of a Life Table</td>
<td>129</td>
</tr>
<tr>
<td>Separation Factors</td>
<td>129</td>
</tr>
<tr>
<td>Calculation of separation factors from statistics</td>
<td>130</td>
</tr>
<tr>
<td>Separation factor for age under 1 year</td>
<td>130</td>
</tr>
<tr>
<td>Separation factors for ages 1 to 4 years</td>
<td>132</td>
</tr>
<tr>
<td>Calculation of separation factors from observed</td>
<td>133</td>
</tr>
<tr>
<td>relationships</td>
<td>133</td>
</tr>
<tr>
<td>From existing life tables</td>
<td>133</td>
</tr>
<tr>
<td>From models and correlations</td>
<td>134</td>
</tr>
<tr>
<td>Calculation of Infant Mortality Rates</td>
<td>135</td>
</tr>
<tr>
<td>Based on reliable statistics</td>
<td>135</td>
</tr>
<tr>
<td>Based on other sources of information</td>
<td>138</td>
</tr>
<tr>
<td>Child Mortality at Ages 1 to 4 Years</td>
<td>138</td>
</tr>
<tr>
<td>Smoothing Death Rates</td>
<td>139</td>
</tr>
<tr>
<td>Life Table</td>
<td>139</td>
</tr>
<tr>
<td>Software</td>
<td>140</td>
</tr>
<tr>
<td>Spreadsheet LTOPDPDTH</td>
<td>140</td>
</tr>
<tr>
<td>Spreadsheet LTMXQXAD</td>
<td>141</td>
</tr>
<tr>
<td>Appendix III-3</td>
<td>143</td>
</tr>
<tr>
<td>Selecting a Model Life Table and Life Expectancy at Birth</td>
<td>143</td>
</tr>
<tr>
<td>from Limited Data</td>
<td>143</td>
</tr>
<tr>
<td>Spreadsheet ADJMX</td>
<td>143</td>
</tr>
<tr>
<td>Spreadsheets LTNTH, LTSTH, and LTWST</td>
<td>144</td>
</tr>
<tr>
<td>Appendix III-4</td>
<td>147</td>
</tr>
<tr>
<td>Stable Population</td>
<td>147</td>
</tr>
<tr>
<td>Appendix III-5</td>
<td>149</td>
</tr>
<tr>
<td>Preston-Coale Technique</td>
<td>149</td>
</tr>
<tr>
<td>Appendix III-6</td>
<td>153</td>
</tr>
<tr>
<td>Growth Balance Mortality Technique</td>
<td>153</td>
</tr>
<tr>
<td>Appendix III-7</td>
<td>157</td>
</tr>
<tr>
<td>Bennett-Horiuchi Mortality Technique</td>
<td>157</td>
</tr>
<tr>
<td>Appendix III-8</td>
<td>159</td>
</tr>
<tr>
<td>Integrated Technique</td>
<td>159</td>
</tr>
<tr>
<td>Appendix III-9</td>
<td>161</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Preston-Bennett Mortality Technique</td>
<td>161</td>
</tr>
<tr>
<td>Appendix III-10</td>
<td>163</td>
</tr>
<tr>
<td>Techniques for Indirect Estimation of Infant and Child Mortality</td>
<td>163</td>
</tr>
<tr>
<td>Appendix III-11</td>
<td>167</td>
</tr>
<tr>
<td>Johnson Technique</td>
<td>167</td>
</tr>
<tr>
<td>Appendix III-12</td>
<td>169</td>
</tr>
<tr>
<td>Orphanhood and Widowhood Techniques</td>
<td>169</td>
</tr>
<tr>
<td>Appendix III-13</td>
<td>175</td>
</tr>
<tr>
<td>Model Life Tables</td>
<td>175</td>
</tr>
<tr>
<td>Appendix III-14</td>
<td>177</td>
</tr>
<tr>
<td>Smoothing with Logits</td>
<td>177</td>
</tr>
<tr>
<td>Spreadsheets LOGITQX and LOGITLX</td>
<td>177</td>
</tr>
<tr>
<td>Program BESTFT</td>
<td>178</td>
</tr>
<tr>
<td>Constructing a Life Table from Fragmentary Data</td>
<td>179</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>180</td>
</tr>
<tr>
<td>Chapter IV</td>
<td>183</td>
</tr>
<tr>
<td>FERTILITY</td>
<td>183</td>
</tr>
<tr>
<td>This Chapter in Brief</td>
<td>183</td>
</tr>
<tr>
<td>Direct estimation</td>
<td>183</td>
</tr>
<tr>
<td>Indirect estimation</td>
<td>183</td>
</tr>
<tr>
<td>Spreadsheets and Methods That Can Be Used for Analyzing Fertility, According to the Available Information</td>
<td>184</td>
</tr>
<tr>
<td>Where to Find the Software</td>
<td>185</td>
</tr>
<tr>
<td>Introduction</td>
<td>186</td>
</tr>
<tr>
<td>Direct Estimation of Fertility</td>
<td>187</td>
</tr>
<tr>
<td>Crude birth rate (CBR)</td>
<td>187</td>
</tr>
<tr>
<td>General fertility rate (GFR)</td>
<td>188</td>
</tr>
<tr>
<td>Age-specific fertility rates (ASFR)</td>
<td>188</td>
</tr>
<tr>
<td>Total fertility rate (TFR)</td>
<td>190</td>
</tr>
<tr>
<td>Gross reproduction rate (GRR)</td>
<td>192</td>
</tr>
<tr>
<td>Net reproduction rate (NRR)</td>
<td>194</td>
</tr>
<tr>
<td>Consistency Among Fertility Rates</td>
<td>195</td>
</tr>
<tr>
<td>Standardization of Crude Birth Rates</td>
<td>196</td>
</tr>
<tr>
<td>Direct standardization</td>
<td>196</td>
</tr>
<tr>
<td>Indirect standardization</td>
<td>197</td>
</tr>
<tr>
<td>Standardization of Total Fertility Rates</td>
<td>198</td>
</tr>
<tr>
<td>Indirect Estimation of Fertility</td>
<td>198</td>
</tr>
<tr>
<td>Techniques based on enumerated population</td>
<td>199</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>----</td>
</tr>
<tr>
<td>Rejuvenation technique</td>
<td>199</td>
</tr>
<tr>
<td>Own-children technique</td>
<td>200</td>
</tr>
<tr>
<td>Thompson technique</td>
<td>202</td>
</tr>
<tr>
<td>Rele technique</td>
<td>204</td>
</tr>
<tr>
<td>Techniques based on special fertility questions</td>
<td>205</td>
</tr>
<tr>
<td>Data problems and adjustments</td>
<td>206</td>
</tr>
<tr>
<td>P/F ratio technique</td>
<td>207</td>
</tr>
<tr>
<td>Brass P1/F1 ratio technique</td>
<td>208</td>
</tr>
<tr>
<td>Arriaga technique</td>
<td>209</td>
</tr>
<tr>
<td>Relational Gompertz Technique</td>
<td>211</td>
</tr>
</tbody>
</table>

Appendix IV-1 ........................................ 213

Relationships Among Fertility Rates .................. 213

Spreadsheet ADJASFR .................................. 213
Spreadsheet TFR-GFR .................................. 214
Spreadsheet CBR-GFR .................................. 215
Spreadsheet CBR-TFR .................................. 216

Appendix IV-2 ........................................ 219

Standardization of the Total Fertility Rate ........ 219

Appendix IV-3 ........................................ 223

Rejuvenation Technique ................................ 223

Appendix IV-4 ........................................ 225

Own-Children Technique ................................ 225

Appendix IV-5 ........................................ 227

Rele Technique ........................................ 227

Appendix IV-6 ........................................ 229

P/F Ratio Technique ................................... 229

Appendix IV-7 ........................................ 233

Arriaga Technique .................................... 233

Appendix IV-8 ........................................ 237

Relational Gompertz Technique ......................... 237

REFERENCES ............................................ 240

Chapter V .............................................. 243
Appendix VIII-6 ......................................................... 367
Projecting the Trend of Fertility ................................. 367
Appendix VIII-7 ......................................................... 369
Projecting the Pattern of Age-Specific Fertility Rates .......... 369
Appendix VIII-8 ......................................................... 371
The Impact of Errors in the Base Population of a Projection .... 371
REFERENCES .......................................................... 376
Chapter IX .............................................................. 377

THE WRITTEN REPORT ............................................... 377
  Introduction ......................................................... 377
  Nontechnical Report ............................................... 377
  Technical Report .................................................. 378
    General technical presentation .................................. 378
    Introduction .................................................... 378
    Data availability ............................................... 378
    Reliability of the data ......................................... 378
    Methods applied ................................................. 378
    Estimates ....................................................... 378
  Component projection of the population ......................... 379
    Projection of the components ................................. 379
    Projection of the population ................................. 379
PREFACE

The main purposes of this manual are to encourage developing countries to make a primary analysis of population information and to make it easier for them to do so.

The U.S. Bureau of the Census, through the International Programs Center (IPC), provides demographic training to population analysts in developing countries. This training is offered as seminars, workshops, and technical assistance for persons interested in doing a demographic analysis of the population. It focuses on: a) the methods and techniques available for analyzing the base data on the demographic situation in a country, and b) the use of microcomputer programs to perform the calculations quickly and accurately.

Both the Bureau of the Census and the sponsoring agencies have recognized the need for a manual containing selected topics presented in the seminars, both as a textbook during the training and as a reference for later use. Until now, the seminars and workshops were offered with a set of references to articles and books that required a solid educational background in the fields of demographic techniques and mathematics. With a view to simplifying the explanation of the demographic techniques and their implementation through microcomputer programs, we undertook the task of producing this manual.

Our goal was to produce a textbook avoiding mathematical derivations of the techniques frequently used in basic demographic analysis. Instead, we wanted to provide a textbook that can be easily understood, given the concepts of the demographic indices and techniques presented, but with references to more detailed sources. The text offers mainly the concepts, while the appendices present a more technical description.

The principal contribution of the manual is not only the easily readable text, but also a set of spreadsheets developed during recent years for analyzing census and survey information. These Population Analysis Spreadsheets were designed and developed primarily by the author, but with significant contributions from his colleagues in the Center.

A set of diskettes containing the spreadsheets is included with the manual and written documentation. The manual and spreadsheets are not copyrighted and may be widely copied and used. Another computer package extensively mentioned in the mortality chapter is the United Nations MORTPAK package. MORTPAK, as well as the United Nations program QFIVE, is distributed separately. Users of this manual can obtain the United Nations programs by writing to the following address.
To request MORTPAK and QFIVE, write to:
Project Coordinator
Computer Software and Support for Population Activities
Department of Technical Cooperation for Development
Room DC2-1570
United Nations
New York, New York 10017

My software associate, Peter D. Johnson, not only contributed some of the spreadsheets, but also undertook the task of standardizing the presentation of the spreadsheets and their conversion from SuperCalc to Lotus 1-2-3. Also, in order to facilitate the use of the spreadsheets, he developed a Lotus program with macros to enhance the user-friendliness of the Lotus spreadsheets. Finally, he wrote the software and documentation of the computer program (RUP) for projecting the population, as presented in Volume II.

My editorial associate, Ellen Jamison, made a substantial contribution by rewriting much of the text from the user’s point of view. She also made useful suggestions regarding format and presentation. Finally, she edited the complete text and made recommendations for improving the documentation of the spreadsheets.

Any errors that remain are my own responsibility.

The manual has two volumes. Volume I includes the technical and theoretical aspects about how to analyze the population information. At the beginning of each chapter an overview of the contents is presented. A reference table indicates the methods and software that can be used according to the available information. Volume II presents the documentation of the Population Analysis Spreadsheets and population projection program. The spreadsheets are presented in alphabetical order.

Eduardo E. Arriaga
November 1994
ACKNOWLEDGEMENTS

This publication was prepared in the International Programs Center (IPC), Population Division, U.S. Bureau of the Census. The author and associates wish to acknowledge the contribution of current and former staff members of IPC to the report: Kathleen S. Short helped the author with the initial version of the spreadsheets; Peter O. Way and Patricia M. Rowe contributed some of the spreadsheets; and Sylvia D. Quick, John M. Reed, Frank B. Hobbs, Arjun Adlakha, Timothy B. Fowler, Thomas McDevitt, and Kevin G. Kinsella made valuable comments for improving the manual and spreadsheets. Donna M. Dove, John R. Gibson, and Peggy Seybolt organized the materials for presentation in the volumes. Other members of IPC also contributed by testing the set of spreadsheets and suggesting improvements. But among all of them, the author and associates thank Barbara Boyle Torrey, former Chief of IPC, for her decision to produce this manual and for her constant encouragement during its preparation. The present chief of IPC, Judith Banister, has reconfirmed the Center's endorsement.

We also recognize the benefits received from comments made by Edward W. Fernandez of the Population Division.

We are grateful to the Office of Population, U.S. Agency for International Development, for its support throughout the various stages of this project, especially to Duff G. Gillespie, Director, and members of the Policy and Evaluation Division: Scott R. Radloff, Chief, Elizabeth Schoenecker and John G. Crowley; also to Judith R. Seltzer, formerly of the same office.

Outside of the U.S. Government, particular thanks are due to Louise Kantrow of the Department of Technical Co-operation for Development of the United Nations, whose suggestions, made when she was at the Policy and Evaluation Division of the Agency for International Development, strengthen the presentation of the text and whose recommendations for the creation of the interface system of the spreadsheets enhance the user friendliness of the package.

Population Analysis with Microcomputers (Volumes I and II) was produced by the Bureau of the Census in collaboration with the International Institute for Vital Registration and Statistics and with financial support from the United Nations Population Fund (UNFPA) and the United States Agency for International Development (USAID).
Chapter I

INTRODUCTION

For adequate planning on the national and regional levels, a nation requires detailed information about the characteristics of the society and about the specific goals of government programs to improve living conditions. Also required is some knowledge about the potential impact and effects of such programs on the society and its development. Such requirements speak for themselves and do not need to be dwelt upon here.

In their quest for social and economic development, developing countries often struggle with information that is incomplete or is not available at the time it is needed. Data are not only required, they are required at the opportune time for use before becoming obsolete.

In most developing countries, the availability of data has improved greatly in recent decades. All countries have expanded and strengthened the capabilities of their statistical offices, including activities related to information on population. By now, all but two countries of the world have conducted at least one population census, and most countries have started to take them regularly. In addition, most nations have begun to take housing, agricultural and industrial censuses as well.

Together with improvements in data collection, the development of computer programs for processing data has enhanced the prompt availability of tabulations appropriate for national planning. The increase in communication among professionals and the availability of technical assistance to those countries lacking needed programs or facilities have accelerated the process of collecting and publishing information. In some cases, the quality of the information still can be improved greatly. In this respect, the area that comes most readily to mind is the needed development and improvement in vital registration systems.

But the availability of information is not the only concern. If data are available in a country but are not analyzed, it is the same as if the data did not exist. The analysis, too, must be timely, as it may rapidly become obsolete in a highly dynamic society. The development of microcomputer programs can accelerate the process of analyzing the data.

The purpose of this manual is to enhance the process of analyzing the available information on population. It includes descriptions of the most frequently used procedures or methods in basic demographic analysis and the corresponding microcomputer programs. Keeping in mind that the manual is intended for use primarily by developing countries with limited information, the array of techniques presented is not exhaustive.

The text is written in such a way that the inexperienced analyst will easily understand the procedures described. Mathematical derivations of the
techniques are not included, but a user with the need to know can consult the references cited. On the other hand, the assumptions that were made in developing the techniques are included, along with the possible biases that result if the data do not meet the conditions assumed in developing the techniques.

The programs referenced in this manual pertain to two packages and one special program. One package is a set of spreadsheets developed at the U. S. Bureau of the Census. This set is called PAS, Population Analysis Spreadsheets. These spreadsheets are used through Lotus, and hence, the user should have access to Lotus 1-2-3 for using the PAS system. The spreadsheets and the special program RUP (a population projection program for two areas) are distributed together with the manual. The other package referenced in this manual is the United Nations MORTPAK set of programs for mortality analysis. The user should obtain this package through the United Nations.

The manual is divided into two volumes. Volume I presents the text of the various topics and includes for each topic, a conceptual explanation of the methods and techniques that can be used for analyzing the population information. Each chapter of the volume has several appendices where further explanation and clarification of the methods and techniques are presented. Volume II is dedicated to the software; microcomputer spreadsheets for analyzing population information are described and documented. This volume includes a system for using the spreadsheets for users who are not familiar with microcomputers. Finally, it presents a population projection program.

**Volume I**

There are nine chapters in volume I. They include the following main topics: age structure, mortality, fertility, migration, distribution of population, urbanization, and population projections.

**Topics Covered**

Following this introduction, there are eight more chapters. All but chapter IX have their own appendices, most of which include a discussion of a particular technique and instructions on how to use the corresponding microcomputer program to perform the calculations involved in the technique. The remaining chapters are described briefly below.

Each chapter starts with a section, *This Chapter in Brief*, in which a summary of the chapter is presented. This section includes a table relating the available information to the software that can be used. The purpose is to provide the user with a quick reference to the methods that can be applied according to the data available. A list of the microcomputer programs used in the chapter is also included.

**Chapter II** discusses the analysis of the age and sex composition of a population. First, it presents some of the indices frequently used to measure age misreporting for each sex and the effect of age misreporting on the sex ratios at each age. Second, it presents smoothing techniques to correct for
age misreporting. References are given to the spreadsheets for analyzing the age and sex distribution of the population and for performing the smoothing.

Chapter III tells how to estimate and analyze mortality. First, it presents some conventional procedures for estimating mortality when data are reliable. Second, it presents some procedures for evaluating the available data and for estimating indirectly the levels of mortality. References to microcomputer programs performing the techniques are given.

Chapter IV tells how to estimate and analyze fertility. As in the previous chapter, it first presents some conventional procedures for estimating fertility when information is reliable. The second part deals with methodologies for estimating fertility indirectly using census or survey information. References to the spreadsheets performing the methods are given.

Chapter V is dedicated to the analysis of migration. Basic techniques for detecting internal migration are presented. Some ideas concerning the estimation of international migration are also given. A reference is given to the spreadsheet for estimating migrants.

Chapter VI suggests how to describe the geographic distribution of the population on a territory. It presents the most frequently used procedures, and gives a reference to the spreadsheet that calculates the related indices.

Chapter VII describes the process of urbanization in a population and the most frequently used indices to measure both the degree of urbanization and the tempo or speed of the urbanization process. A reference is given to the spreadsheet that calculates these indices.

Chapter VIII deals with methods for projecting the population. Most of the chapter is dedicated to the population projection procedure known as the component method, by age and sex. It also tells how to project each of the projection components: mortality, fertility, and migration. The last part of the chapter describes some procedures for projecting the total population (not by age) using mathematical functions. While the text discusses the concepts involved in the projection procedures, the appendices present the formulas. References to the spreadsheets frequently used in the estimation process are given.

Chapter IX provides suggestions about the format and content of a report describing the basic demographic characteristics of a country, once the analysis has been completed. This chapter has no appendices.

Volume II

A set of 45 spreadsheets, a menu for using them, and a program for projecting the population of two areas simultaneously are included in volume II. The spreadsheets and the menu for using them constitute the Population Analysis Spreadsheet (PAS) system.
The set of spreadsheets was developed in Lotus. Users familiar with this system can use the spreadsheets directly with it. For those who are not familiar with Lotus, the PAS system facilitates use of the spreadsheets (see appendix I-1).

Each spreadsheet has its documentation, where the user will find a brief description of the procedure used in the spreadsheet, instructions on how to use it, and where to incorporate the data. The documentation, tables, and graphics presented in volume II are exactly the same as what the user will see on the computer screen when using the spreadsheets.

The population projection program for projecting the population of two areas of the country simultaneously is called RUP. A brief explanation of the method used, the data requirements, and the results obtained, are presented in chapter VIII of volume I. The methodology used, the documentation, and the required computer specifications are presented in volume II.

How to Use This Manual

The text of this manual has been prepared for persons without a strong background in demographic techniques. It is intended for use by persons who are applying such techniques for the first time (though it may serve as a review for experienced users as well). For this reason, no mathematical derivation of the techniques is given, but the concepts and microcomputer programs to perform the techniques are provided.

The recommended steps for conducting demographic analysis using the manual are as follows:

1. Study and understand the concepts involved in each of the techniques presented for the demographic variables to be analyzed.

2. Take note of the information required for each technique that will permit the estimation of the levels of these variables, and of the assumptions underlying the methodology of each one.

3. Review the country's data situation compared with the data needed for each technique, as well as the country's conditions as related to the underlying assumptions.

4. Select the techniques that can be used with the available data and whose assumptions are compatible with the country's situation.

5. Use the microcomputer programs corresponding to the selected techniques to carry out the analysis.

6. Interpret the results.

7. Prepare a report presenting the analysis, including a summary discussion of the data and methods used.
Some Guidelines for Using the Spreadsheets

A set of recommendations is presented here to facilitate the proper use of the spreadsheets for users who are not familiar with Lotus. These recommendations will help to avoid inadvertently modifying the calculations in the spreadsheets, and to avoid errors.

(1) Before using each spreadsheet for the first time, read the instructions. The manual presents the instructions for each spreadsheet in the documentation of the programs. In addition, during the use of the spreadsheet with the computer, you can obtain the instructions by pressing the key Alt and at the same time the letter H (referred to in this manual as Alt-H).

(2) All cells in the spreadsheets are protected except those where data or titles are to be entered as input. DO NOT UNPROTECT any protected cells. The purpose of the protection is to avoid destroying the formulas. If you overwrite a cell containing a formula, the formula will be destroyed and the spreadsheet cannot be used. If this happens by mistake, reload the original spreadsheet and enter your data again.

(3) Do not modify the original set of spreadsheets. If you install the spreadsheets on a hard disk, do not save a used spreadsheet in the same file; save it under another name. If you are using the original diskettes, apply a "write protect" tab to the edge of each diskette so no files can be stored on them. The purpose of the disk protection is to reduce the chance of accidentally destroying or modifying the original spreadsheets.

(4) While you are becoming familiar with the spreadsheets, it will be useful to save the ones you create in a certain way. Before saving a spreadsheet containing your data, press Alt-H, and the instructions will appear on the screen. Then save the spreadsheet. In this way, the next time you load the same spreadsheet, the instructions will automatically appear on the screen. Later, when you are more familiar with the programs, it is not necessary to perform the Alt-H function before saving.

(5) Check the accuracy of your data input before calculating, as input errors are frequently made. In some spreadsheets using data on population, deaths, or births, the data are summed automatically. In these cases, check if the totals from your sources are equal to those calculated by the spreadsheets.

(6) Try to avoid using the command move (in Lotus) while using the spreadsheets. The command move would modify the reference to certain cells and probably the spreadsheet would not work properly. Similarly, when using the command copy for transferring data from one spreadsheet to another, use the copy value command to avoid copying formulas that do not pertain to the spreadsheet.
(7) The data input as viewed on the screen of some spreadsheets includes commas to separate thousands and millions (for instance, 23,435,128). DO NOT TYPE THE COMMAS (enter the figure as 23435128). The spreadsheet is programmed so the commas will automatically appear on the screen. If you type commas, the input will be interpreted as text, and the spreadsheet will not work. However, decimal points must be entered when needed.

(8) During data entry, be careful not to confuse the letter "o" with the cipher "0" nor the letter "l" with the number "1." If you mistakenly enter letters into a cell where numbers are expected, you will get errors or unpredictable results.

(9) Dates to which the estimates pertain usually are not the same as the reference date of the input data. For instance, if infant mortality is estimated based on 1987 census information on children ever born and children surviving provided by females in the age group 25 to 29 years, the date of the estimate is not 1987, but a few years earlier.
The Population Analysis Spreadsheet (PAS) system is a set of Lotus 1-2-3 spreadsheets that have been created to help in analyzing demographic data and preparing for population projections. The PAS system consists of two parts: (a) the set of spreadsheets for demographic analysis and (b) the PAS menu, a special Lotus spreadsheet with macros for facilitating the use of the spreadsheets.

The PAS spreadsheets can be used directly with Lotus 1-2-3 in the customary manner; hence, those who are familiar with Lotus can use the spreadsheets without the PAS macro menu. For those who are not familiar with Lotus, the set of macros of the PAS menu will facilitate the use of the spreadsheets. Following the instructions given in this appendix, the user will be able to retrieve, use and save the demographic spreadsheets. The major advantage of the PAS menu is that it cautions the user during the process of saving a file to avoid destroying the original spreadsheets. It also prevents the user from exiting a spreadsheet without saving it.

<table>
<thead>
<tr>
<th>Spreadsheet</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJAGE</td>
<td>Adjusts an age/sex distribution to a new total.</td>
</tr>
<tr>
<td>ADJASFR</td>
<td>Proportionally adjusts ASFR's to get total births.</td>
</tr>
<tr>
<td>ADJMIX</td>
<td>Proportionally adjusts $m_x$ values to get total deaths.</td>
</tr>
<tr>
<td>AGEINT</td>
<td>Interpolates between two age/sex distributions.</td>
</tr>
<tr>
<td>AGESEX</td>
<td>Calculates the U.N. age/sex accuracy index.</td>
</tr>
<tr>
<td>AGESMTH</td>
<td>Smooths an age/sex distribution using several methods.</td>
</tr>
<tr>
<td>ARFE-2</td>
<td>Arriaga fertility method--data for 2 dates.</td>
</tr>
<tr>
<td>ARFE-3</td>
<td>Arriaga fertility method--data for 3 dates.</td>
</tr>
<tr>
<td>ASFRPATT</td>
<td>Estimates ASFR's with a given TFR.</td>
</tr>
<tr>
<td>BASEPOP</td>
<td>Estimates and smooths a base population consistent with past mortality and fertility.</td>
</tr>
<tr>
<td>BPSTRNG</td>
<td>Estimates and strongly smooths a base population consistent with past mortality and fertility.</td>
</tr>
<tr>
<td>Command</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>BTHSRV</td>
<td>Estimates IMR based on the survival of infants born in last year.</td>
</tr>
<tr>
<td>CBR-GFR</td>
<td>Estimates CBR and GFR from TFR and female population by age.</td>
</tr>
<tr>
<td>CBR-TFR</td>
<td>Estimates CBR and TFR from GFR and female population by age.</td>
</tr>
<tr>
<td>CSRMIG</td>
<td>Census survival ratio method for estimating net internal migration.</td>
</tr>
<tr>
<td>CTBL32</td>
<td>Contingency table adjustment of a 2-variable distribution.</td>
</tr>
<tr>
<td>EOLGST</td>
<td>Extrapolates life expectancy by sex using a logistic function.</td>
</tr>
<tr>
<td>FITLGSTC</td>
<td>Fits a logistic to a set of 3 or multiple of 3 points.</td>
</tr>
<tr>
<td>GRBAL</td>
<td>Brass growth balance equation method.</td>
</tr>
<tr>
<td>GRPOP-YB</td>
<td>Graphs birth cohorts from 2 or 3 censuses.</td>
</tr>
<tr>
<td>INTPLTF</td>
<td>Interpolates a female life table with a given life expectancy based on two life tables.</td>
</tr>
<tr>
<td>INTPLTM</td>
<td>Interpolates a male life table with a given life expectancy based on two life tables.</td>
</tr>
<tr>
<td>LOGISTIC</td>
<td>Extrapolates an index by fitting a logistic function.</td>
</tr>
<tr>
<td>LOGITLX</td>
<td>Logit transformation of $l_x$ values.</td>
</tr>
<tr>
<td>LOGITQX</td>
<td>Logit transformation of $q_x$ values.</td>
</tr>
<tr>
<td>LTMXQXAD</td>
<td>Life table based on adjusted $m_x$ or $q_x$ values.</td>
</tr>
<tr>
<td>LTNTH</td>
<td>Selects a Coale-Demeny north model life table by sex based on a crude death rate and population by age and sex.</td>
</tr>
<tr>
<td>LTPOPDTH</td>
<td>Life table based on population and deaths by age.</td>
</tr>
<tr>
<td>LTSTH</td>
<td>Selects a Coale-Demeny south model life table by sex based on a crude death rate and population by age and sex.</td>
</tr>
<tr>
<td>LTWST</td>
<td>Selects a Coale-Demeny west model life table by sex based on a crude death rate and population by age and sex.</td>
</tr>
<tr>
<td>MOVEPOP</td>
<td>Moves an age/sex distribution to a new date.</td>
</tr>
<tr>
<td>OPAG</td>
<td>Estimates the population in the open-ended age group.</td>
</tr>
<tr>
<td>PFRATIO</td>
<td>Estimates fertility using the P/F ratio method.</td>
</tr>
</tbody>
</table>
PREBEN  Estimates mortality using the Preston-Bennett method.
PRECOA  Estimates completeness of information on deaths using Preston-Coale method.
PYRAMID Creates a population pyramid.
RELEFERT Rele method for estimating fertility.
REL-GMPZ Relational Gompertz method for estimating fertility.
REVCBR  Estimates the CBR based on reverse survival of the population under age 15.
SINGAGE Analyzes the population by single years of age.
SP      Constructs a stable population.
TFRLGST Extrapolates the TFR by fitting a logistic function.
TFRSINE Extrapolates the TFR by fitting a sine function.
TFR-GFR  Estimates TFR and GFR from CBR and female population in reproductive ages.
URBINDEX Urbanization indices.
Chapter II

AGE AND SEX COMPOSITION

This Chapter in Brief

The source of data on the age and sex composition of a population is usually a population census. Although all modern censuses collect information on age and sex of the population, the data often contain errors because some people do not know their true age and others do not report their age accurately.

The first step in evaluating the age and sex composition is to make a graphical comparison of the age structure of the population from several consecutive censuses. For this purpose, see the software programs GRPOP-YB and PYRAMID.

There are several indices for evaluating the age and sex composition. Most of these indices rely on an expected pattern reflecting the distribution of a population without migration and in which mortality and fertility have changed in only one direction. The principal indices are:

(1) sex ratios;
(2) age ratios; and
(3) indices for detecting digit preference in age reporting.

The recommended software programs for calculating these indices are: AGESEX, AGESMTH, and SINGAGE.

Because of data errors, population age structures must be adjusted with techniques for smoothing the irregularities. The techniques for smoothing age structures are:

(1) Slight smoothing, which gently modifies irregularities in the age structure:

   (a) methods which keep the enumerated population within each 10-year age group; and

   (b) methods which modify the enumerated population totals.

(2) Strong smoothing, which modifies most irregularities, some of which may represent actual facts instead of errors.

The recommended software for smoothing is called AGESMTH.
<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet Procedure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population age structure by 5-year age groups</td>
<td>PYRAMID</td>
<td>Makes a population pyramid.</td>
</tr>
<tr>
<td></td>
<td>ADJAGE</td>
<td>Adjusts the population age structure (for each sex or both combined) to a new population total.</td>
</tr>
<tr>
<td></td>
<td>AGESEX</td>
<td>Computes age and sex indices.</td>
</tr>
<tr>
<td></td>
<td>AGESMTH</td>
<td>Smooths population age distributions.</td>
</tr>
<tr>
<td>Population age structure by 5-year age groups for more than one date</td>
<td>GRPOP-YB</td>
<td>Graphs the population age structure by birth cohorts.</td>
</tr>
<tr>
<td></td>
<td>AGEINT</td>
<td>Interpolates the age structure between two dates.</td>
</tr>
<tr>
<td>Population age structure by single years of age</td>
<td>SINGAGE</td>
<td>Analyzes digit preference of age reporting.</td>
</tr>
<tr>
<td>Population age distribution, age specific death and fertility rates (5-year age groups), and migrants</td>
<td>MOVEPOP</td>
<td>Moves the population by age to a new date.</td>
</tr>
<tr>
<td>Open-ended age groups that are too broad</td>
<td>OPAG</td>
<td>Distributes the population of the open-ended age group by 5-year age groups up to 80 years.</td>
</tr>
<tr>
<td></td>
<td>CTBL32</td>
<td>Same as OPAG, but simultaneously for several open-ended age groups.</td>
</tr>
</tbody>
</table>
Where to Find the Software

The following spreadsheets for analyzing the age structure of the population can be found in volume II.

Bureau of the Census spreadsheets:

- ADJAGE
- AGEINT
- AGSEX
- AGESMTH
- CTBL32

Introduction

The distribution of a population by age and sex is one of the most basic types of information needed in planning for the future. Any analysis of educational requirements, military needs, labor force projections, family composition, retirement, migration, or voting practices, for example, would not be complete without considering information on age. Age is a crucial component in demographic analysis as well. The study of mortality and fertility without considering age will permit only a partial understanding of these phenomena.

Given the importance of the age structure with respect to social and economic characteristics, it is imperative that the information on the population age and sex structure be as accurate as possible. The following sections discuss the graphic presentation, evaluation, and adjustment of data on age.

Age Structure as a View of the Past

A population's age structure may be considered as a map of its demographic history. Persons of the same age constitute a cohort of people who were born during the same year (or period); they have been exposed to similar historical facts and conditions in the nation. The age structure of the whole population at a given moment may be viewed as an aggregation of cohorts born in different years. A graphic representation of the age structure of the population, such as an "age pyramid," shows the different surviving cohorts of people of each sex in the country.
The age pyramid illustrated in figure II-1-A represents a population in which fertility did not fluctuate significantly during the past, mortality had a typical trend, migration was probably insignificant, and age reporting seems to be accurate. Figure II-1-B, on the other hand, represents a current population in a European country where the age structure has been affected by the impact of mortality and fertility changes during the war. Finally, a country in which a large number of recent male immigrants have altered the age structure is illustrated in figure II-1-C. (See appendix II-1 and the software program PYRAMID, in volume II.)

Checking Census Data for Consistency

Comparing historical population age distributions for a country helps in analyzing data consistency. This may be accomplished graphically by plotting the population in the various age groups by year of birth. Such a figure reveals past trends of fertility, migration, age misreporting, and even errors in census enumeration. The graph cannot disentangle actual demographic history from statistical errors, but it can indicate in general whether the age distribution is acceptable as reported or if it needs some adjustment.

Figure II-2 presents information on the age structure by year of birth for three consecutive decennial censuses in a sample country whose population has not been exposed to migration. Each line represents a census, the top one the earliest census and each subsequent lower one a later census. The graph shows how the population born in the same period of years and enumerated in each successive census is reduced through time. In this case, the spaces between the lines should represent the reduction of each cohort due to persons who died during the intercensal period. In other cases, part of the difference may reflect errors in the data or different levels of census completeness.

The area of the figure labeled "A" shows the survivors of persons born during the period 1930 to 1945 as enumerated in the 1960, 1970, and 1980 censuses. Strictly parallel lines would indicate reliable census information, in spite of the unexpected indentation found in the three censuses. The repetition of the indentation in successive censuses in the same cohorts supports a hypothesis of lower fertility levels during a particular decade compared to earlier and later decades, unless the cohorts were exposed to unusually high mortality from birth to the time of the first census. An increase of mortality in specific ages would likely reflect an epidemic that would be well known in the country. The absence of knowledge of any such epidemic leaves only the possibility of a reduction of birth rates during the late 1930's and early 1940's.
Figure 11-1-A. Population Age Distribution Corresponding to a Condition of Small Past Changes in Fertility and Mortality

Figure 11-1-B. Population Age Distribution Corresponding to Conditions of Past Changes in Fertility and Mortality

Figure 11-1-C. Population Age Distribution Corresponding to a Condition of Recent Significant Male Immigration
Large migratory movements can also be detected from the plotting of population census data by age and year of birth. In figure II-3, the data from points A to B indicate a cohort of persons born between 1926 and 1965 who were 15 to 54 years of age at the time of the 1980 census; this cohort is larger than it was when enumerated in the 1975 census, when it was 10 to 49 years of age. Unless there is evidence of a large undercount of the population ages 10 to 49 years in the 1975 census, or a large overcount of ages 15 to 54 years in the 1980 census, it may be safely concluded that migrants are responsible for the increase of the cohort during the intercensal period. For constructing such population age structure figures by year of birth, see the software program GRPOP-YB in volume II.

**Detecting age misreporting**

Population data in developing countries are often subject to age misreporting. Irregularities in the age distribution produced by respondents' incorrect age declaration can be detected in graphical cohort analysis. As illustrated in figure II-4, age misreporting may be suggested by the repetition of a similar age pattern for different cohorts (as opposed to the parallelism expected for the same cohort). In this comparison of data from two censuses, it seems as if the age pattern of the population at the two dates has been shifted, as the two age distributions appear to be "out of line."
Figure II-3. Census Population by Year of Birth Indicating Immigration

Figure II-4. Census Population by Year of Birth Indicating Age Misreporting
For example, the shortage of people in ages 10 to 19 years in the 1960 census (points C and D of figure II-4) is repeated at the same ages in the 1970 census (points A and B). The same situation appears in the older ages; the pattern at ages 55 to 69 years in 1960 (points H, I, and J) is repeated at the same ages in the 1970 census (points E, F, and G). This could be due to an attraction of age 60 (irregularities due to digit preference will be discussed below).

Although migration might at first be suspected as the cause of this distortion of the age pattern, the age of migrants as well as the direction of the migration movement would be extremely unusual, making it difficult to accept that migration caused the distortion. In this case, the overall repetitive pattern should be interpreted as errors in the data, which thus would need some adjustment before using them.

Age misreporting in the process of enumerating the population in a census can come from two sources. One source is the respondent, who either willfully misreports his or her age or gives an approximation if the true age is unknown. The other source of error is the interviewer who estimates the age of a respondent who does not know his or her age. In either case, the result of this age-guessing process is that ages are often rounded to end in the digits 0 and 5.

**Digit preference**

Irregularities in reporting single years of age can be detected by using indices or graphs. There are several frequently used indices for detecting digit preference: Myers (1940), Whipple (U.S. Bureau of the Census, 1971), Bachi (1951, 1953), Carrier (1959), and Ramachandran (U.S. Bureau of the Census, 1971). These indices not only provide an overall idea of the extent of age misreporting but also indicate the preference for certain ending age digits. (See appendix II-2 and the software program SINGAGE in volume II.) However, the analysis can be done graphically by constructing a typical population pyramid by single years of age. The single-age pyramid presented in figure II-5 shows a notorious age misreporting in the country's population in certain ages (such as 30, 40, 50, and 60 years). Information on age containing such errors cannot be accepted as reported, and so adjustments would be required in this case.
Age ratios

Age ratios for 5-year age groups have also been proposed as indices for detecting possible age misreporting in populations where fertility has not fluctuated greatly during the past and where international migration has not been significant. Under such demographic conditions, age ratios are expected to be similar throughout the age distribution, and all of them should be rather close to a value of 100. Age ratios are calculated by dividing the population of a specific 5-year age group by the average population of the two adjacent 5-year age groups, times 100 (United Nations, 1952). The larger the fluctuations of these ratios and the larger their departure from 100, the greater the probability of errors in the data (except in populations exposed to international migration). An example is presented in table II-1 and figure II-6, where age ratios are calculated using the same populations as in figures II-1-A and II-4. The formulas and an example of their application are presented in appendix II-2, and the software programs AGESEX and AGESMTH in volume II.
Figure II-6. Age Ratios Pertaining to Populations Shown in Figures II-1-A and II-4

Note: Each ratio represents the population in a given age group divided by the average population of the two adjacent age groups, times 100.

**Sex ratios**

Sex ratios are another analytical tool and are easily calculated by dividing the male population in a given age (or age group) by the female population in the same age (or age group), times 100. As in the case of age ratios, the larger the abrupt departure of this ratio from values close to 100, the larger the possibility of errors in the data. The level of the sex ratios depends on the number of male and female births and on the mortality of the population. All populations have more male than female births, and so the sex ratio at the early ages is expected to be slightly over 100. However, since mortality is usually higher for males than females, the sex ratio is reduced continuously up to the oldest ages. There are exceptions to this pattern: (a) sometimes female mortality is higher than male mortality; and (b) a population may have a significant number of migrants of just one sex.

To make an illustrative comparison, sex ratios are shown for the populations presented in figures II-1-A and II-4 (see figure II-7). While the sex ratios of the population of figure II-1-A follow a pattern with a smooth trend indicating not only good quality data but also no external influences, those pertaining to the population of figure II-4 have strong fluctuations resulting from errors in the data (see figure II-7 and table II-1). See calculations in appendix II-2, and the software programs AGESEX and AGESMTH in volume II.
Figure II-7. Sex Ratios Pertaining to Populations Shown in Figures II-1-A and II-4

Note: Each ratio refers to the number of males per 100 females in each age group.

Age-sex accuracy index

In the early 1950's, the United Nations (1952 and 1955) analyzed in detail the problems of errors in reporting in a population. This institution suggested a joint accuracy index to summarize the values of the age and sex ratios.

(1) The index of sex-ratio score (SRS) was defined as:

*The mean difference between sex ratios for the successive age groups, averaged irrespective of sign.*

(2) The index of age-ratio score (ARS) was defined as:

*The mean deviation of the age ratios from 100 percent, also irrespective of sign.*

The age-ratio score is calculated independently for males (ARSM) and females (ARSF).
### Table II-1. Population Age and Sex Distribution and Age and Sex Ratios: Ghana, 1960.

<table>
<thead>
<tr>
<th>Age</th>
<th>Population</th>
<th>Age ratios</th>
<th>Sex</th>
<th>Age ratio</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Total</td>
<td>3,400,270</td>
<td>3,326,545</td>
<td>102.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-4</td>
<td>642,367</td>
<td>654,258</td>
<td></td>
<td>98.2</td>
<td></td>
</tr>
<tr>
<td>5-9</td>
<td>515,520</td>
<td>503,070</td>
<td>103.1</td>
<td>102.9</td>
<td>102.5</td>
</tr>
<tr>
<td>10-14</td>
<td>357,831</td>
<td>323,460</td>
<td>90.5</td>
<td>84.2</td>
<td>110.6</td>
</tr>
<tr>
<td>15-19</td>
<td>275,542</td>
<td>265,534</td>
<td>88.0</td>
<td>82.2</td>
<td>103.8</td>
</tr>
<tr>
<td>20-24</td>
<td>268,336</td>
<td>322,576</td>
<td>96.8</td>
<td>112.8</td>
<td>83.2</td>
</tr>
<tr>
<td>25-29</td>
<td>278,601</td>
<td>306,329</td>
<td>100.1</td>
<td>107.8</td>
<td>90.9</td>
</tr>
<tr>
<td>30-34</td>
<td>242,515</td>
<td>245,883</td>
<td>101.7</td>
<td>101.3</td>
<td>99.6</td>
</tr>
<tr>
<td>35-39</td>
<td>198,231</td>
<td>179,182</td>
<td>97.1</td>
<td>91.5</td>
<td>110.6</td>
</tr>
<tr>
<td>40-44</td>
<td>165,937</td>
<td>145,572</td>
<td>103.4</td>
<td>106.0</td>
<td>114.0</td>
</tr>
<tr>
<td>45-49</td>
<td>122,756</td>
<td>95,590</td>
<td>93.5</td>
<td>84.1</td>
<td>128.4</td>
</tr>
<tr>
<td>50-54</td>
<td>96,775</td>
<td>81,715</td>
<td>106.3</td>
<td>113.5</td>
<td>118.4</td>
</tr>
<tr>
<td>55-59</td>
<td>59,307</td>
<td>48,412</td>
<td>74.0</td>
<td>71.0</td>
<td>122.5</td>
</tr>
<tr>
<td>60-64</td>
<td>65,467</td>
<td>56,572</td>
<td>138.4</td>
<td>141.8</td>
<td>116.3</td>
</tr>
<tr>
<td>65-69</td>
<td>32,377</td>
<td>28,581</td>
<td>69.4</td>
<td>70.3</td>
<td>113.3</td>
</tr>
<tr>
<td>70-74</td>
<td>29,796</td>
<td>26,733</td>
<td>122.7</td>
<td>123.3</td>
<td>111.5</td>
</tr>
<tr>
<td>75-79</td>
<td>16,183</td>
<td>16,778</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80+</td>
<td>34,729</td>
<td>30,300</td>
<td></td>
<td>114.6</td>
<td></td>
</tr>
</tbody>
</table>

(a) Sex ratio score = 8.3 Average of the sex ratios, column (5).
(b) Male age ratio score = 12.5 Average of the age ratios, column (3).
(c) Female age ratio score = 16.1 Average of the age ratios, column (4).
(d) Accuracy index = 53.7 \(3 \times (a) + (b) + (c)\)

Notes: Age ratios (columns 3 and 4) are 100 times the ratio of each age group divided by the average of adjacent age groups.
Sex ratios (column 5) are 100 times the male population divided by the female population in each age group.
Age ratio deviation is the difference of each age ratio from 100 regardless of the sign.
Difference of sex ratios is the difference between two consecutive sex ratios regardless of the sign.
(3) Based on empirical relationships between the sex-ratio scores and the age-ratio scores, the following index was defined as the joint score (JS) or age-sex accuracy index:

\[
JS = 3xSRS + ARSM + ARSF
\]

Based on empirical analysis of the age and sex declaration in censuses from different developed and developing countries, the United Nations suggested that the age and sex structure of a population will be (a) accurate if the joint score index is under 20, (b) inaccurate if the joint score index is between 20 and 40, and (c) highly inaccurate if the index value is over 40.

These indices are useful mainly in international or historical comparative analyses. Historical series of indices indicate whether the quality of the population age and sex reporting is improving or deteriorating.

Although these indices quantify the quality of age and sex information from censuses and surveys, a graphical analysis as mentioned before should always be conducted.

The calculation of these indices is performed by the Bureau of the Census spreadsheet AGESEX, presented in volume II.

Correcting for Age Misreporting

Once indices or graphs indicate that the age structure of the population is not correct, a decision should be made about whether or not the age structure should be adjusted. Smoothing techniques have frequently been used for correcting data for age misreporting. Most of these techniques involve the application of a formula or function to the original data. They may be classified into two main categories: (a) techniques which accept the population in each 10-year age group and separate it into two 5-year age groups without modifying the total population size; and (b) techniques which smooth the 5-year age groups and modify slightly (either increasing or decreasing) the population being smoothed. Both procedures are discussed below (see appendix II-2 for smoothing formulas).

**Smoothing without modifying the totals**

The various formulas that accept the enumerated population in each 10-year age group give rather similar results. The main difference among them is in whether or not they smooth the first and last 10-year age groups in the distribution. The Carrier-Farrag and Karup-King-Newton (Carrier and Farrag, 1959) formulas do not separate the first or last 10-year age groups, while the Arriaga (1968) formula does.

The Carrier-Farrag technique is based on the assumption that the relationship of a 5-year age group to its constituent 10-year age group is an average of similar relationships in three consecutive 10-year age groups. The Karup-King-Newton formula assumes a quadratic
relationship among each three consecutive 10-year age groups, while the Arriaga formula assumes that a second degree polynomial passes by the midpoint of each three consecutive 10-year age groups and then integrates a 5-year age group. The formulas are presented in appendix II-2.

All these formulas give rather similar results, and thus it is difficult to point out advantages or disadvantages of each one in relation to the others, although the simplicity of the Karup-King-Newton formula is attractive when computers are not available. For comparative purposes, table II-2 presents these smoothing techniques applied to the age distribution of the same population considered in table II-1. See appendix II-2, and the software program AGESMTH in volume II.


<table>
<thead>
<tr>
<th>Age</th>
<th>Carrier Farrag</th>
<th>Karup-King Newton</th>
<th>Arriaga</th>
<th>United Nations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Total</td>
<td>2,161,675</td>
<td>2,097,406</td>
<td>2,161,675</td>
<td>2,097,406</td>
</tr>
<tr>
<td>0-4</td>
<td>662,761</td>
<td>675,049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-9</td>
<td>495,126</td>
<td>482,279</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td>346,290</td>
<td>316,905</td>
<td>354,871</td>
<td>327,523</td>
</tr>
<tr>
<td>15-19</td>
<td>287,083</td>
<td>272,089</td>
<td>278,502</td>
<td>261,471</td>
</tr>
<tr>
<td>20-24</td>
<td>285,855</td>
<td>237,266</td>
<td>285,508</td>
<td>324,698</td>
</tr>
<tr>
<td>25-29</td>
<td>261,082</td>
<td>301,639</td>
<td>261,429</td>
<td>304,207</td>
</tr>
<tr>
<td>30-34</td>
<td>237,937</td>
<td>237,876</td>
<td>236,513</td>
<td>236,766</td>
</tr>
<tr>
<td>35-39</td>
<td>202,809</td>
<td>187,189</td>
<td>204,233</td>
<td>188,299</td>
</tr>
<tr>
<td>40-44</td>
<td>162,973</td>
<td>138,294</td>
<td>162,138</td>
<td>139,015</td>
</tr>
<tr>
<td>45-49</td>
<td>125,720</td>
<td>102,868</td>
<td>126,555</td>
<td>102,147</td>
</tr>
<tr>
<td>50-54</td>
<td>88,730</td>
<td>73,673</td>
<td>90,094</td>
<td>74,939</td>
</tr>
<tr>
<td>55-59</td>
<td>67,352</td>
<td>56,454</td>
<td>65,988</td>
<td>55,188</td>
</tr>
<tr>
<td>60-64</td>
<td>55,187</td>
<td>47,474</td>
<td>54,803</td>
<td>47,115</td>
</tr>
<tr>
<td>65-69</td>
<td>40,657</td>
<td>35,679</td>
<td>41,941</td>
<td>36,838</td>
</tr>
<tr>
<td>70-74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75-79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Smoothing modifying the totals**

The most frequently used formula is that presented long ago by the United Nations (Carrier-Farrag, 1959). It is calculated based on five 5-year age groups, for smoothing the central group in each instance. The difference between the total adjusted and unadjusted populations is usually small. If the enumerated totals are desired, proportional adjustment to such totals is advised for those smoothed age groups (see table II-2). See appendix II-2, and the software program AGESMTH in volume II.

**Selection of a smoothing technique**

The smoothing of age distributions by the above formulas usually gives acceptable results. However, these formulas produce only a light smoothing, and if age misreporting is severe, light smoothing probably will not be sufficient.

For instance, age distributions such as those presented in figures II-2 and II-3 do not have significant irregularities with respect to age misreporting. The population presented in figure II-2 has a cohort of persons born during the 1930's and 1940's which is smaller than expected as a consequence of lower fertility during that period. Hence, the indentation detected in all censuses (area A of figure II-2) should not be smoothed out because it is an actual fact. Therefore, if it is decided that the population needs some smoothing due to slight irregularities in other ages, the smoothing procedure should be light.

A similar conclusion can be drawn from a population age distribution such as the one presented in figure II-3, where the 1980 census reflects some intercensal immigration in relation to the 1975 population.

However, the population age distribution shown in figure II-4 is a different case. This population needs some smoothing because most of the irregularities in the age structure are due to age misreporting and digit preference for some ages. The irregularities are so large, that if the age structure is smoothed with any of the formulas presented above, the results still will not be free of irregularities (see figure II-8). Most of the above formulas accept the population in each 10-year age group and, hence, if age is misreported by more than 10 years, the problem is not solved by such formulas. It appears that part of the population in age group 20 to 29 years pertains to younger ages. Consequently, the solution is to smooth the age structure with a formula that will modify the population in each 10-year age group. In other words, the population data require a "stronger" smoothing procedure.
**Strong smoothing**

An easily applied strong smoothing procedure is as follows: (a) first smooth the 10-year age groups; (b) adjust the results to the census population in smoothed ages; and (c) separate the smoothed 10-year age groups into 5-year age groups using any of the formulas presented above.

In the case of figure II-4, the 1960 census population needs a strong smoothing. The following steps were followed (table II-3):

1. The census population was combined into 10-year age groups.
2. The 10-year age groups from age 10 to age 69 years were smoothed by averaging three consecutive 10-year age groups with specific weights. For instance, the smoothed age group 30 to 39 years was obtained as
one-fourth of age group 20 to 29, plus one-half of age group 30 to 39, plus one-fourth of age group 40 to 49. This procedure was followed for all 10-year age groups between ages 10 and 70 years (see table II-3, columns 3 and 4).

(3) Since the total population of these smoothed age groups was greater than the census population of the same ages, the smoothed age groups were adjusted proportionally to the census totals (see table II-3, columns 5 and 6).

(4) Finally, the smoothed and adjusted 10-year age groups were subdivided into 5-year age groups using Arriaga's (1968) formula (see table II-3, columns 7 and 8; appendix II-2; and the software program AGESMTH in volume II).

While figure II-8 compares the census and the "light" smoothed populations, figure II-9 shows the census and the "strong" smoothed populations. Both figures illustrate the differences between the procedures for smoothing a population. Other characteristics of the census population and the population smoothed by the two procedures can be seen as sex ratios in table II-4 and figures II-10 and II-11, and as age ratios in table II-5 and figures II-12 and II-13.

There is no generalized solution for all populations. The smoothing technique to be used will depend on the errors in the age and sex distributions, and so the age structure must be analyzed before deciding whether the smoothing should be strong or light. This brings us back to the first steps of this section: make a graph of the age and sex distributions before making any decision about whether or not smoothing is required and which formula or technique would be appropriate for the particular country's situation. Comparisons among successive censuses and a knowledge of past trends of mortality, fertility, and migration will help in appraising the accuracy of the reported age and sex structure of the population.

<table>
<thead>
<tr>
<th>Age</th>
<th>Census population</th>
<th>Smoothed population 10-69</th>
<th>Adjusted population 10-69</th>
<th>Adjusted population 5-year age groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>0-4</td>
<td>1,157,887</td>
<td>1,157,328</td>
<td>1,157,887</td>
<td>1,157,328</td>
</tr>
<tr>
<td>5-9</td>
<td>633,373</td>
<td>588,994</td>
<td>742,893</td>
<td>741,055</td>
</tr>
<tr>
<td>10-14</td>
<td>546,937</td>
<td>628,905</td>
<td>541,998</td>
<td>557,967</td>
</tr>
<tr>
<td>15-19</td>
<td>440,746</td>
<td>425,065</td>
<td>429,281</td>
<td>430,049</td>
</tr>
<tr>
<td>20-24</td>
<td>288,693</td>
<td>241,162</td>
<td>293,554</td>
<td>259,379</td>
</tr>
<tr>
<td>25-29</td>
<td>156,082</td>
<td>130,127</td>
<td>174,175</td>
<td>146,142</td>
</tr>
<tr>
<td>30-34</td>
<td>95,844</td>
<td>83,153</td>
<td>98,437</td>
<td>84,486</td>
</tr>
<tr>
<td>35-39</td>
<td>45,979</td>
<td>41,511</td>
<td>45,979</td>
<td>41,511</td>
</tr>
<tr>
<td>40-44</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
<tr>
<td>45-49</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
<tr>
<td>50-54</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
<tr>
<td>55-59</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
<tr>
<td>60-64</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
<tr>
<td>65-69</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
<tr>
<td>70-74</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
<tr>
<td>75-79</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
<tr>
<td>80+</td>
<td>34,729</td>
<td>30,300</td>
<td>34,729</td>
<td>30,300</td>
</tr>
</tbody>
</table>

Total 10-69 2,161,675 2,097,406 2,161,675 2,097,406 2,161,675 2,097,406

Notes:
- a) Columns (3) and (4): each 10-year age group in columns (1) and (2) is smoothed as the following average: \( S(o) = \frac{(T(-1)+2xT(o)+T(1))}{4} \)
- b) Columns (5) and (6): the smoothed 10-year age groups in columns (3) and (4) are adjusted to census totals for ages 10 to 69.
- c) Columns (7) and (8): age groups columns (5) and (6) are separated into 5-year age groups by using Arriaga’s formula.
Figure II-9. Actual and Smoothed Population with Strong Smoothing

Figure II-10. Sex Ratios: Actual Population and with Light Smoothing

Note: Each ratio refers to the number of males per 100 females in each age group.
Figure II-11. Sex Ratios: Actual Population and with Strong Smoothing

Note: Each ratio refers to the number of males per 100 females in each age group.

Figure II-12. Age Ratios: Actual Population and with Light Smoothing

Note: Each ratio represents the population in a given age group divided by the average population of the two adjacent age groups, times 100.
Figure II-13. Age Ratios: Actual Population and with Strong Smoothing

Note: Each ratio represents the population in a given age group divided by the average population of the two adjacent age groups, times 100.
### Table II-4. Sex Ratios of the Actual and Smoothed Populations: Ghana, 1960.

<table>
<thead>
<tr>
<th>Age</th>
<th>Census</th>
<th>Carrier-King</th>
<th>Farrag</th>
<th>Arriaga</th>
<th>United Nations</th>
<th>Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
</tr>
<tr>
<td>0-4</td>
<td>98.2</td>
<td>98.2</td>
<td>99.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-9</td>
<td>102.5</td>
<td>102.7</td>
<td>100.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td>110.6</td>
<td>110.3</td>
<td>108.3</td>
<td>109.8</td>
<td>109.4</td>
<td>101.4</td>
</tr>
<tr>
<td>15-19</td>
<td>103.8</td>
<td>105.5</td>
<td>106.5</td>
<td>104.9</td>
<td>100.8</td>
<td>100.5</td>
</tr>
<tr>
<td>20-24</td>
<td>83.2</td>
<td>87.3</td>
<td>87.9</td>
<td>86.7</td>
<td>87.0</td>
<td>96.1</td>
</tr>
<tr>
<td>25-29</td>
<td>90.9</td>
<td>86.6</td>
<td>85.9</td>
<td>87.3</td>
<td>89.0</td>
<td>96.2</td>
</tr>
<tr>
<td>30-34</td>
<td>98.6</td>
<td>100.0</td>
<td>99.9</td>
<td>100.5</td>
<td>99.1</td>
<td>98.9</td>
</tr>
<tr>
<td>35-39</td>
<td>110.6</td>
<td>108.3</td>
<td>108.5</td>
<td>107.7</td>
<td>108.8</td>
<td>102.6</td>
</tr>
<tr>
<td>40-44</td>
<td>114.0</td>
<td>117.8</td>
<td>116.6</td>
<td>117.6</td>
<td>116.9</td>
<td>111.9</td>
</tr>
<tr>
<td>45-49</td>
<td>128.4</td>
<td>122.2</td>
<td>125.9</td>
<td>122.5</td>
<td>123.4</td>
<td>116.8</td>
</tr>
<tr>
<td>50-54</td>
<td>118.4</td>
<td>120.4</td>
<td>120.2</td>
<td>120.4</td>
<td>122.9</td>
<td>119.5</td>
</tr>
<tr>
<td>55-59</td>
<td>122.5</td>
<td>119.3</td>
<td>119.6</td>
<td>119.4</td>
<td>119.2</td>
<td>120.8</td>
</tr>
<tr>
<td>60-64</td>
<td>116.3</td>
<td>116.2</td>
<td>116.3</td>
<td>117.4</td>
<td>118.2</td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>113.3</td>
<td>114.0</td>
<td>113.9</td>
<td>114.1</td>
<td>113.4</td>
<td>116.3</td>
</tr>
<tr>
<td>70-74</td>
<td>111.5</td>
<td>111.9</td>
<td></td>
<td></td>
<td>113.0</td>
<td></td>
</tr>
<tr>
<td>75-79</td>
<td>109.5</td>
<td>109.0</td>
<td></td>
<td></td>
<td>107.4</td>
<td></td>
</tr>
</tbody>
</table>

Source: Column 2 calculated from Table II-1; columns 3-6 calculated from Table II-2; and column 7 calculated from Table II-3, columns 7 and 8.

### Table II-5. Age Ratios of the Actual and Smoothed Populations: Ghana, 1960.

<table>
<thead>
<tr>
<th>Age</th>
<th>Census population</th>
<th>Carrier-Farrag</th>
<th>K.King-Newton</th>
<th>Arriaga</th>
<th>United Nations</th>
<th>Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male Female</td>
<td>Male Female</td>
<td>Male Female</td>
<td>Male Female</td>
<td>Male Female</td>
<td>Male Female</td>
</tr>
<tr>
<td>0-4</td>
<td>103.1 102.9</td>
<td>98.2 97.4</td>
<td>98.9 98.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-9</td>
<td>90.5 84.2</td>
<td>88.3 83.2</td>
<td>93.4 92.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-14</td>
<td>88.0 82.2</td>
<td>90.8 84.5</td>
<td>87.0 80.2</td>
<td>91.1</td>
<td>85.1</td>
<td>96.1</td>
</tr>
<tr>
<td>15-19</td>
<td>112.8 104.3</td>
<td>114.1 100.2</td>
<td>114.8 104.2</td>
<td>115.1</td>
<td>97.5</td>
<td>106.1</td>
</tr>
<tr>
<td>20-24</td>
<td>107.8 109.1</td>
<td>106.7 100.2</td>
<td>105.7 104.2</td>
<td>110.2</td>
<td>97.5</td>
<td>98.8</td>
</tr>
<tr>
<td>25-29</td>
<td>101.3 102.6</td>
<td>97.3 101.6</td>
<td>96.1 102.3</td>
<td>96.9</td>
<td>103.0</td>
<td>102.1</td>
</tr>
<tr>
<td>30-34</td>
<td>101.5 99.5</td>
<td>102.5 100.2</td>
<td>101.8 101.0</td>
<td>98.9</td>
<td>96.0</td>
<td>101.1</td>
</tr>
<tr>
<td>35-39</td>
<td>91.2 95.5</td>
<td>101.2 102.5</td>
<td>101.8 101.0</td>
<td>98.9</td>
<td>96.0</td>
<td>101.1</td>
</tr>
<tr>
<td>40-44</td>
<td>106.0 99.2</td>
<td>98.0 95.7</td>
<td>97.9 94.1</td>
<td>99.5</td>
<td>97.0</td>
<td>98.5</td>
</tr>
<tr>
<td>45-49</td>
<td>106.0 99.9</td>
<td>97.1 100.3</td>
<td>95.5 101.4</td>
<td>98.2</td>
<td>99.3</td>
<td>100.6</td>
</tr>
<tr>
<td>50-54</td>
<td>113.5 91.9</td>
<td>92.5 93.6</td>
<td>93.6 91.1</td>
<td>90.4</td>
<td>94.3</td>
<td>97.9</td>
</tr>
<tr>
<td>55-59</td>
<td>71.0 93.6</td>
<td>93.2 91.1</td>
<td>90.4 94.3</td>
<td>93.7</td>
<td>92.2</td>
<td>97.9</td>
</tr>
<tr>
<td>60-64</td>
<td>141.8 138.4</td>
<td>102.2 103.1</td>
<td>102.4 103.3</td>
<td>100.4</td>
<td>101.4</td>
<td>102.6</td>
</tr>
<tr>
<td>65-69</td>
<td>70.3 69.4</td>
<td>99.0 99.4</td>
<td>97.5 97.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70-74</td>
<td>123.3 122.7</td>
<td>98.5 99.1</td>
<td>96.5 96.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75-79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Column 2 calculated from Table II-1; columns 3-6 calculated from Table II-2; and column 7 calculated from Table II-3, columns 7 and 8.
Other age structure considerations

A few additional aspects of the age structure analysis should be considered. They are the open-ended age group, the adjustment of a population age structure to a particular total, and the shifting of a population from one reference date to another.

Open-ended age group. The open-ended age group obviously cannot be smoothed. But sometimes the terminal age group is "too young," for instance, 65 years and over. When it is desirable to break down a part of the open-ended age group into 5-year age groups, the problem arises as to how this should be done.

Two possibilities are presented here for breaking down the open-ended age group into 5-year age groups for cases where the age of the open-ended age group is younger than 80+ years. One of the procedures is to use a contingency table, while the other is to use stable population theory. The first procedure requires that preliminary distributions of the open-ended age group be available. The second requires an estimate of the mortality level of the population.

The contingency table is recommended for cases where the problem exists at the state or provincial level, and the age distribution for the whole country has 5-year age groups at least up to age 79 years. The distribution for the whole country is used as a control for breaking down the open-ended age groups of the states or provinces, which start at younger ages. It also requires a preliminary 5-year age group distribution for each state's open-ended age group (for example the state's distributions from other years). These preliminary distributions have little impact on the final age distributions for each state. See appendix II-3 for a discussion, and volume II for the software program CTBL32.

The procedure for breaking down the open-ended age groups into 5-year age groups using stable population theory can be used for any population where the level of mortality can be "guessed." The required level of mortality has little impact on the estimated 5-year age distribution of the open-ended age group. See appendix II-4 for a discussion, and volume II for the software program OPAG.

Population adjustment. The adjustment of a population age structure to another population total may be made proportionally. This process is used in the spreadsheet ADJAGE developed by the Bureau of the Census and presented in volume II. This spreadsheet allows the user to provide a new total for both sexes combined (and each sex is calculated proportionally to the actual data) or a new total for each sex separately.
Population interpolation. Population information may be needed for dates other than the census date. If the desired date is between census dates, an estimate of the population can be obtained by interpolating between the two censuses. For instance, if population censuses are available for April 7, 1981 and July 25, 1990, interpolated populations can be obtained for any desired date in between.

Two interpolation processes can be applied: (a) using the same cohort in each census; (b) using the same age groups in each census. Cohort interpolation is advisable if the information is reliable. This interpolation is recommended if data are available by single ages for populations that have actual and significant age fluctuations (produced by fertility, migration or mortality). In addition, information on annual intercensal births is needed for interpolating the population of persons born during the intercensal period, who were enumerated only in the last census.

Interpolation using the same age groups in each census is frequently used in developing countries, since information on single ages is unreliable, and the age structure usually does not have actual fluctuations. The procedure is acceptable, particularly when the interpolation date is close to the census dates.

Interpolation using the same age groups in each census can be made linearly or exponentially. If the desired date is close to one of the census dates, both possibilities give similar results. But, if the interpolation date is several years distant from the census dates, linear interpolation will produce higher values than exponential. In most of the cases, the exponential interpolation will be more suitable than the linear. The formulas for both methods of interpolation are presented in appendix II-6.

The Bureau of the Census has developed a spreadsheet (AGEINT) that performs both the linear and the exponential interpolation. The documentation of this spreadsheet can be seen in volume II.

Population shifting. Another procedure for adjusting the population is to shift or move the population from a given date (for instance, a census date) to another date (for instance, midyear). If the total population for the desired date is available, the procedure explained above can be used. Another possibility is to use the levels of mortality, fertility and migration (if any) of the population to estimate the population growth rate, and then use the growth rate to calculate the population total at the desired date. Once the new total is calculated, the age structure can be adjusted proportionately to the new total. See volume II for the software program MOVEPOP.
Appendix II-1

Graphing Population Age Structures

Age Pyramid

Age pyramids are usually constructed by considering the male and female population separately, by 5-year age groups. Age pyramids can represent the actual population, a percent distribution of the population of each sex with 100 percent equal to the total of the given sex, or a percent distribution of the population of each sex with 100 percent equal to the total population (both sexes combined). The selection from these alternatives depends on the purpose of the graph. Software programs are provided to produce the three types of graphs.

Population by Year of Birth

The purpose of graphing the population by year of birth rather than by age groups is to relate certain irregularities of the age structure to historical facts that may have affected the age distribution. The usefulness of such a graph is greater if more than one census is included. In this case, cohorts are followed easily on the vertical axis of the graph. For easy detection of irregularities, the actual population figures are plotted on a semi-logarithmic scale.

Software

There are two programs developed by the Bureau of the Census to make the graphs of the population age structure: PYRAMID and GRPOP-YB. The documentation for PYRAMID and GRPOP-YB is presented in volume II. In addition, the program SINGAGE gives a graph with the distribution of population by single years of age; documentation of this program also is presented in volume II.
Appendix II-2

Analyzing and Smoothing the Age Distribution of a Population

As mentioned in the text, the reported age distribution of a population may contain errors that can be detected and reduced by some procedures. To detect the errors, indices for age-digit preference, sex ratios, and age ratios are used (see text sections on software SINGAGE and AGESEX). To reduce the errors, smoothing procedures are used. The formulas and examples of the procedures described in the text are presented here. The smoothing techniques are classified into two categories: those which keep the population of each 10-year age group, and those which modify it (see text section on software AGESMTH).

Analyzing the Age Structure

The analysis of the age and sex structure can be conducted by single ages or by age groups. Single ages are sometimes analyzed in order to determine whether the population age reporting was affected by digit preference. In most populations, more people are reported than expected in ages ending in 0 and 5 because of preference for those digits.

Although a graph showing the age structure by single ages is a good instrument for judging the age reporting of the population, there are also some indices for analyzing the preference for some digits. Whipple, Myers, Bachi, Carrier, and Ramachandran have developed such indices (U.S. Bureau of the Census, 1971).

The Bureau of the Census has developed a spreadsheet called SINGAGE which, in addition to graphing the population by single years of age, also calculates three indices: Whipple’s, Myers’, and Bachi’s. A brief explanation of these indices is presented here.

Whipple’s index (U.S. Bureau of the Census, 1971) detects a preference for ages ending in 0, 5, or both. The concept is to take an age interval (with a multiple of 10, such as age 40 or 50 years) starting with ages ending in digits other than 0 or 5. The next step is to sum the population in ages ending in 0 (or in ages ending in 5), multiply the sum by 10, and divide the result by the population of the age interval under study. If the index is for ages 0 and 5 combined, the sum of the population in ages ending in those digits is multiplied by 5 and divided by the population in the whole age interval. If age reporting is correct, this index should fluctuate slightly around 1. The higher the value of the index, the higher the preference for digits 0 or 5.

The Myers (1940) and Bachi (1951, 1953) indices are similar, although the magnitude of the Myers index is almost double that of the Bachi index. Both indices give a magnitude of the excess or deficit of people in ages ending in any of the 10 digits. Such excesses or deficits are expressed as percentages. The larger the value of the indices, the larger the preference...
for certain digits. Values close to 0 would indicate excellent age reporting in censuses.

There are some other indices used for 5-year age groups which also analyze the age and sex structure of the population. The most commonly used indices are the age and sex ratios (United Nations, 1952)

**Age ratios**

An age ratio is defined as:

\[
5\text{AR}_x = 100 \frac{5P_x}{1/2 (5P_{x-5} + 5P_{x+5})}
\]

Where:

- \(5\text{AR}_x\) represents the age ratio for ages \(x\) to \(x+4\); and
- \(5P_x\) represents the population at ages \(x\) to \(x+4\).

The following example illustrates this formula for males ages 10 to 14 years (data are from table II-1);

<table>
<thead>
<tr>
<th>Age</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-9</td>
<td>515,520</td>
</tr>
<tr>
<td>10-14</td>
<td>357,831</td>
</tr>
<tr>
<td>15-19</td>
<td>275,542</td>
</tr>
</tbody>
</table>

\[
5\text{AR}_{10} = \frac{100 \times 357,831}{1/2 (515,520 + 275,542)} = 90.5.
\]

**Sex ratios**

A sex ratio at a given age is defined as:

\[
5\text{SR}_x = 100 \frac{5\text{MP}_x}{5\text{FP}_x}
\]

Where:

- \(5\text{SR}_x\) represents the sex ratio at ages \(x\) to \(x+4\); and
- \(5\text{MP}_x\) and \(5\text{FP}_x\) represent the male and female populations, respectively, at ages \(x\) to \(x+4\).

The following example illustrates this formula for ages 10 to 14 years (data are from table II-1):
Population

<table>
<thead>
<tr>
<th>Age</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-14</td>
<td>357,831</td>
<td>323,460</td>
</tr>
</tbody>
</table>

\[ s_{SR_{10}} = 100 \times \frac{357,831}{323,460} = 110.6. \]

**Smoothing the Age Distribution**

This section provides the formulas corresponding to the various demographic techniques described in this chapter for those that do not appear in the text where the methods are described. All the formulas are included in the spreadsheet called AGESMTH.

**Smoothing without modifying the totals of each 10-year age group**

**Carrier-Farrag formula.** These formulas (Carrier and Farrag, 1959) are as follows:

\[ s_{P_{x+5}} = \frac{\underline{10P_x}}{[1 + \left(\frac{\underline{10P_{x-10}}}{\underline{10P_{x+10}}}\right)^{1/4}] } \] and

\[ s_{P_x} = \underline{10P_x} - s_{P_{x+5}} \]

Where:

- \(s_{P_{x+5}}\) represents the population at ages \(x+5\) to \(x+9\);
- \(\underline{10P_x}\) represents the population at ages \(x\) to \(x+9\); and
- \(s_{P_x}\) represents the population at ages \(x\) to \(x+4\).

The following example illustrates this formula for males ages 25 to 29 years (data are from table II-3):

<table>
<thead>
<tr>
<th>Age</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-19</td>
<td>633,373</td>
</tr>
<tr>
<td>20-29</td>
<td>546,937</td>
</tr>
<tr>
<td>30-39</td>
<td>440,746</td>
</tr>
</tbody>
</table>

\[ s_{P_{25}} = \frac{546,937}{[1 + (633,373 / 440,746)^{1/4}] } = 261,082 \]

Once a 5-year age group is calculated from a 10-year age group, the other 5-year age group is found by subtraction. For example:

\[ s_{P_{20}} = \underline{10P_{20}} - s_{P_{25}} = 546,937 - 261,082 = 285,855 \]

**Karup-King-Newton formula.** These formulas (Carrier and Farrag, 1959) are as follows:
\[
5P_x = \frac{1}{2} 10P_x + \frac{1}{16} (10P_{x-10} - 10P_{x+10}) \quad \text{and}
\]

\[
5P_{x+5} = 10P_x - 5P_x
\]

Where the symbols are as above (see Carrier-Farrag formula).

The following example illustrates this formula for males ages 20 to 24 years (data are from table II-3, same as above):

\[
5P_{20} = \frac{1}{2} 546,937 + \frac{1}{16} (633,373 - 440,746) = 285,508
\]

The age group 25 to 29 years is found by subtraction:

\[
5P_{25} = 261,429 = 546,937 - 285,508
\]

**Arriaga formula.** When the 10-year age group to be separated is the central group of three, the following formulas (Arriaga, 1968) are used:

\[
5P_{x+5} = (-10P_{x-10} + 11 \cdot 10P_x + 2 \cdot 10P_{x+10}) / 24 \quad \text{and}
\]

\[
5P_x = 10P_x - 5P_{x+5}
\]

Where:

- \(5P_{x+5}\) is the population ages \(x+5\) to \(x+9\);
- \(10P_x\) is the population ages \(x\) to \(x+9\); and
- \(5P_x\) represents the population at ages \(x\) to \(x+4\).

For example, from the information on 10-year age groups (from table II-3) as presented above:

\[
5P_{25} = 261,018 = (-633,373 + 11 \cdot 546,937 + 2 \cdot 440,746) / 24
\]

and

\[
5P_{20} = 285,919 = 546,937 - 261,018.
\]

When the 10-year age group to be separated is an extreme age group (the youngest or the oldest), the formulas are different. For the youngest age group, the following formulas are used:

\[
5P_{x+5} = (8 \cdot 10P_x + 5 \cdot 10P_{x+10} - 10P_{x+20}) / 24 \quad \text{and}
\]

\[
5P_x = 10P_x - 5P_{x+5}
\]

40
Where the symbols are as above.

For the oldest age group, the coefficients are reversed:

\[ sP_x = \left(-\frac{10P_{x-20}}{10} + 5 \frac{P_{x-10}}{10} + 8 \frac{P_x}{10}\right) / 24 \quad \text{and} \]

\[ sP_{x+5} = \frac{10P_x}{10} - \frac{5P_x}{10} \]

For example, given the following information,

<table>
<thead>
<tr>
<th>Age</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-9</td>
<td>1,157,887</td>
</tr>
<tr>
<td>10-19</td>
<td>633,373</td>
</tr>
<tr>
<td>20-29</td>
<td>546,937</td>
</tr>
<tr>
<td>50-59</td>
<td>156,082</td>
</tr>
<tr>
<td>60-69</td>
<td>95,844</td>
</tr>
<tr>
<td>70-79</td>
<td>45,979</td>
</tr>
</tbody>
</table>

the age group 0 to 9 years is separated as follows:

\[ sP_3 = 495,126 = (8 \times 1,157,887 + 5 \times 633,373 - 546,937) / 24 \]

and

\[ sP_0 = 662,761 = 1,157,887 - 495,126. \]

The age group 70 to 79 years is separated as follows:

\[ sP_{70} = 28,790 = (-156,082 + 5 \times 95,844 + 8 \times 45,979) / 24 \]

and

\[ sP_{75} = 17,189 = 45,979 - 28,790. \]

**Smoothing Modifying the Totals of Each 10-Year Age Group**

**United Nations formula.** United Nations (Carrier and Farrag, 1959) developed the following formula:

\[
5P'_x = \frac{1}{16} \left(-5P_{x-10} + 4 \frac{5P_{x-5}}{5} + 10 \frac{P_x}{10} + 4 \frac{5P_{x+5}}{5} - 5P_{x+10}\right)
\]

Where:

\[ 5P'_x \] represents the smoothed population ages \( x \) to \( x+4 \).

The following example illustrates this formula for males ages 25 to 29 years (data are from table II-1):
\[
\frac{1}{16} \left( - 275,542 + 4 \times 268,336 + 10 \times 278,601 + 4 \times 242,515 - 198,231 \right) = 272,228
\]

**Strong smoothing formula.** If a strong smoothing is desired (Arriaga, 1968), this can be achieved with the following formula:

\[
10P'_x = \left( \frac{10P_{x-10} + 2 \times 10P_x + 10P_{x+10}}{4} \right)
\]

Where:

\(10P'_x\) represents the smoothed population ages \(x\) to \(x+9\).

For example, based on the information above (see Arriaga formula):

\[
10P'_{10} = \left( \frac{1,157,887 + 2 \times 633,373 + 546,937}{4} \right) = 742,893
\]

After proportionally adjusting the smoothed 10-year age groups to the census population in the smoothed ages, the 10-year age groups can be subdivided into 5-year age groups using any of the available formulas.

**Software**

The Bureau of the Census has developed three spreadsheets for analyzing the accuracy of age reporting of a population and for smoothing the age structure. The documentation of these spreadsheets is presented in volume II. These spreadsheets are:

**SINGAGE.** This spreadsheet estimates some indices for analyzing the single-year age declaration of a population. It calculates the indices of Myers, Whipple, and Bachi.

**AGESEX.** This spreadsheet calculates sex ratios and age ratios in a population age distribution.

**AGESMTH.** This spreadsheet smooths the population by 5-year age groups and provides sex and age ratios of the smoothed and unsmoothed populations.
Contingency tables are used for adjusting two variable distributions to specific totals of each of the two variables. The procedure can be used with any two variables. For example, chapter II suggests it be used to break down the open-ended age group, while chapter VIII suggests it be used for adjusting regional population projections to the projection for the whole country. Contingency tables can also be applied to characteristics of educational attainment, labor force, income, languages, religion, and other variables.

**Description**

This technique adjusts information in a table to a set of desired marginal totals (totals of rows and/or columns). For instance, data classified by two characteristics may need to be adjusted to certain totals different from the sum of each column and row. This technique modifies the original data in such a way that after the calculation the sums of columns and rows correspond to the desired totals.

The spreadsheet is designed for using as marginals: (a) the age distribution of the total country or area, located in the first column; and (b) the total population for each subpopulation, located in the first row of data. Instead of age groups, any other variable can be used. The user should remember that the values given as input in the first column will be retained as the proper sum of each of the rows that need adjustment. The totals will be the controls for the sum of each row, and the sum of the column will be the control for the whole population.

On the other hand, the first row of totals does not need to add to the total of the whole population (sum of the first column). If such partial totals (those of the first row) do not add to the whole country, they are adjusted proportionally to the total for the whole country. In contrast, if the totals of the first row add to the total for the whole country, such partial totals are kept.

**Data Required**

1. A preliminary estimate of the data distribution to be adjusted.
2. Desired marginal totals to which the data are to be adjusted.

**Assumptions**

The technique assumes that the original distribution of the information can be adjusted to obtain an estimate of the actual distribution by successive proportional adjustments of rows and columns through an iterative process.
**Procedure**

The original distribution of data is adjusted to the desired marginal totals as follows:

1. Each row of data is adjusted proportionally to the desired total of the row.
2. Each column of data is adjusted proportionally to the desired total of the column.
3. The first two steps are repeated until the data agree with the desired marginal totals of rows and columns.

The procedure can be used for estimating two variable distributions given the marginal totals and a preliminary estimate of the distribution. The preliminary estimate can be a distribution pertaining to another year or to another population.

**Advantages**

The technique provides an estimated distribution of data that matches any set of desired marginal totals. It is useful as a final step in adjusting population projections by regions or states to the total population of a country.

**Limitations**

Requires an initial distribution of the data by the characteristics of the variable.

**Software**

The Bureau of the Census has developed a spreadsheet to apply this procedure. It is called CTBL32, and its documentation is presented in volume II.
Appendix II-4

Estimating the Age Structure of the Open-Ended Age Group

In most cases, the 5-year age distribution of a population is tabulated up to rather old ages and hence the open-ended age group starts at an acceptable age (for example, 80 years and over). However, in some age distributions, the open-ended age is rather young, such as 65 years and over or 70 years and over. Although the latter situation is observed more frequently at the state or provincial level, it may sometimes occur at the national level as well.

In cases where the open-ended age group for the whole country begins at age 80 years or above while those pertaining to states begin at younger ages, the age distributions of the open-ended age group for states can be estimated using a contingency table as described in appendix II-3. There are various options for obtaining the required initial age distribution of the open-ended age group. It can be assumed that: (a) each state distribution is the same as that for the whole country; (b) the age distribution available for each state is the same as the distribution for another year; or (c) a stable population age distribution or even a life table population distribution can be used. The results based on such estimates for each state would be acceptable because the population for the whole country is used as a control.

Given that in developing countries the 5-year age distribution of the population in older ages is not greatly affected by past changes in fertility and/or mortality, or even by migration, a stable population distribution can be used to distribute an open-ended age group into 5-year age groups. The stable population procedure is presented in the spreadsheet OPAG.

Description

The spreadsheet OPAG estimates the population by 5-year age groups for an open-ended age group beginning at ages younger than 80 years and over based on stable population theory (Dublin and Lotka, 1925; Lotka, 1934, 1937, 1939; United Nations, 1968).

Data Required

(1) A distribution of the population by 5-year age groups from age 45 years up to the beginning of the open-ended age group.

(2) The population of the open-ended age group, which must not begin at an age younger than 65 years.

(3) An estimate of life expectancies at birth and at age 80 years. Errors in these estimates have little impact on the estimated age distribution of the open-ended age group. If available, the \( L_x \) function for older ages can be given. Otherwise, the life table
function will be estimated based on the given life expectancy at birth and the Coale-Demeny West model life tables.

Assumptions

This procedure assumes that the population at ages 45 years and above has characteristics similar to those of a stable population (for information on stable population, see appendix III-4 of the next chapter).

Procedure

(1) An estimate of the intrinsic growth rate is calculated based on two 10-year age groups of the population: 45 to 54 years and 55 to 64 years. The formula used is:

\[ r = \frac{\ln(10L_{x+10} / 10L_x) - (\ln 10P_x / 10L_x)}{10} \]

Where \( 10P_x \) and \( 10L_x \) represent the reported population and the life table function at ages \( x \) to \( x+10 \) years, respectively.

(2) A preliminary 5-year age distribution of the open-ended age group is calculated as:

\[ n^{P_x} = n^{L_x} \exp[-r(x+n/2)] \]

for age groups under 80 years of age.

The open-ended age group 80 years and over is calculated using the following formula (Coale-Demeny, 1968):

\[ P_{80+} = L_{80+} \cdot \exp(-r[80 + 0.6E(80) + 0.92]) \]

Where the new symbol \( E(80) \) represents the life expectancy at age 80.

(3) The population of the open-ended age group is distributed proportionally to the population distribution estimated in steps (1) and (2).

Advantages

The procedure requires information only on the population ages 45 years and over, by age, and an estimate of life expectancy at birth and at age 80 years. Errors in these estimates have little impact on the estimation of the 5-year age group distribution of the open-ended age group.
Limitations

(1) Migration and age misreporting at ages 45 to 64 years may affect the estimation process.

(2) The procedure assumes no migration in the ages being estimated.

Software

The Bureau of the Census has developed a spreadsheet called OPAG for estimating the 5-year age distribution of the open-ended age group. See documentation in volume II.
Appendix II-5

Adjusting the Age Structure of the Population to a New Total Population

Description

Population age distributions frequently have to be adjusted to new population totals. For example, once a census population has been analyzed for possible errors and any errors have been corrected, the population total may differ from the accepted total.

This spreadsheet takes a population age distribution and proportionally adjusts it to a desired population total. The adjustment can be made independently to given population totals for each sex, or to a total for both sexes combined. The latter procedure adjusts the total for each sex proportionally to the given total for both sexes combined.

Data Required

(1) A population distribution by 5-year age groups, either for each sex separately or for both sexes combined.

(2) A desired population total, either for each sex separately or for both sexes combined.

Assumption

The procedure assumes that the relative change in the population of each age group is the same as the change in the total population.

Procedure

The population of each age group is multiplied by the ratio of the desired population total to the total of the population age distribution.

Advantage

The procedure can be used to obtain either independent percent distributions of the population of each sex (each sex totals 100 percent) or a percent distribution by age and sex with the total of both sexes combined equal to 100 percent.

Limitations

The procedure adjusts the population of all age groups by the same percentage in spite of the fact that the relative growth in each age group is not the same in a real population.
Software

The Bureau of the Census has developed the spreadsheet ADJAGE, which performs the calculations of this procedure. For documentation of ADJAGE, see volume II.
Appendix II-6

Interpolation Between Two Population Age Structures

Description

The analysis of a population may require an interpolation between two population age structures, for example, when the population age structure is desired for a date between censuses. The linear or exponential interpolation performed by this spreadsheet is made using the population in the same age groups at the two dates.

Data Required

(1) The population age structure for two dates.

(2) An indication of whether linear or exponential interpolation is desired.

Assumptions

The procedure assumes that the average annual change in each population age group has been constant during the intercensal period.

Procedure

(1) The linear interpolation is performed as follows:

\[ P_s = kP_i + (1-k)P_j \]

Where:

- \( P_i, P_s, \) and \( P_j \) represent the population of each age group at dates \( i, s, \) and \( j; \)
- \( s \) is the date for which the interpolation is desired (it must be chronologically between dates \( i \) and \( j); \) and
- \( k \) is a constant for all ages groups, calculated as:
  \[ k = (j-s)/(j-i) \]

(2) The exponential interpolation is performed as follows:

\[ P_s = P_i \exp(hr) \]
Where:

\[ P_1 \text{ and } P_8 \text{ are the same as above; and} \]

\[ h \text{ and } r \text{ are calculated as:} \]

\[ h = s - i \]

\[ r = \frac{\ln(P_j / P_i)}{j-i} \]

**Advantages**

The procedure does not require data on births or deaths, and it offers reasonable results if the interpolation is made for dates close to the census dates.

**Limitations**

For countries that have had migration or large fluctuations of fertility, the interpolation may give unrealistic results for specific population age groups.

**Software**

The spreadsheet AGEINT developed by the Bureau of the Census calculates the interpolations. Its documentation is presented in volume II.
Appendix II-7

Moving the Population to a Desired Date

Description

This spreadsheet was designed for situations where interpolation procedures are not appropriate. The process of interpolating populations by age, as explained in appendix II-6, assumes that the average change of the population in a specific age group during the intercensal period is the same in all years of the time interval. If mortality is rapidly changing, however, the population change in each age group may not be the same at the beginning and end of the intercensal period. In this situation, an interpolation may produce a larger or smaller population than actually was the case at a particular date. In addition, if only one population census exists, the interpolation process cannot be applied.

The spreadsheet MOVEPOP has another approach to such problems by estimating the population growth rate for the census date and using it to move the census total population to another desired date. Then it proportionally distributes the estimated total population by age and sex based on the distribution of the census population.

Data Required

(1) The census population by age and sex.
(2) Sex- and age-specific death rates for the census date.
(3) Age-specific fertility rates for the census date.
(4) The annual net number of migrants for the census year.

Assumptions

The program assumes that the population in each age group changes at the same growth rate as the one estimated for the whole population.

Procedure

The census population and death rates by age and sex, as well as the age-specific fertility rates, are used to estimate the annual number of deaths and births. The annual natural increase and the annual number of migrants are used to estimate the annual growth rate of the population. The growth rate is used to estimate the total population for a desired date. The estimated total population is proportionally distributed according to the age and sex composition of the census population.
Advantages

The procedure takes into account mortality, fertility and migration for calculating the annual growth rate of the population at the census date.

Limitations

The procedure assumes that the population in all age groups changes at the same pace as the total population. If there is a large number of migrants concentrated in particular ages, this assumption may not be valid.

Software

The Bureau of the Census has developed the spreadsheet MOVEPOP for making the calculations of this procedure. Its documentation is presented in volume II.
REFERENCES


Chapter III
MORTALITY

This Chapter in Brief

The purpose of this chapter is to measure the level of mortality in a population. The methods for measuring mortality depend not only on the quality but also on the detail of the information available. If data are reliable, mortality can be estimated directly. If data are not reliable, then specific techniques are required to estimate mortality indirectly.

Direct estimation

Reliable information on population and registered deaths is used to measure the level of mortality. The indices for this measurement are:

(1) crude death rates (with possibilities for direct or indirect standardization); and

(2) construction of life tables to obtain life expectancies at birth and other useful mortality functions.

Life table construction is not difficult, and there are several microcomputer programs which perform the calculations. Special attention should be given to calculating the infant mortality rate and to closing the life table, specifically the open-ended age group. The computer programs available for constructing a life table from reliable information are: Bureau of the Census spreadsheets LTPOPDTH, ADJMX, and LTMXQXAD; and the United Nations program LIFTB in the MORTPAK package.

If the available information is limited, model life tables can be constructed. Programs for constructing model life tables are Bureau of the Census spreadsheets LTNTH, LTSTH and LTWST, and the United Nations program MATCH in the MORTPAK package.

Indirect estimation

There are several techniques for estimating mortality in cases where death statistics are incomplete or nonexistent.

(1) Techniques evaluating unreliable information on deaths are performed by the following programs: Bureau of the Census spreadsheets PRECOA and GRBAL, and the United Nations program BENHR in the MORTPAK package.

(2) Techniques using only age distributions of the population are performed by the following programs: Bureau of the Census
spreadsheet PREBEN and the United Nations program PRESTO in the MORTPAK package.

(3) Techniques using special questions related to mortality from censuses or surveys are performed by several programs:

(a) For estimating infant and child mortality, Bureau of the Census spreadsheet BTHSRV and the United Nations program CEBCS in the MORTPAK package.

(b) For estimating adult mortality, the United Nations programs ORPHAN and WIDOW in the MORTPAK package.

**Smoothing mortality rates**

There are several programs for smoothing mortality rates. Some of them, in addition to smoothing the rates, construct a life table using the smoothed data. These are the Bureau of the Census spreadsheets LTPOPDTH, LOGITLX, and LOGITQX. The United Nations program BESTFT in the MORTPAK package smooths the mortality rates but does not provide a life table. However, the results can be used in conjunction with the program LIFTB of the same package to obtain a life table.

**Spreadsheets, United Nations Programs, and Methods That Can Be Used for Analyzing Mortality, According to the Available Information**

<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population age structure and a crude death rate</td>
<td>LTNTH</td>
<td>Selects a Coale-Demeny north model life table that will reproduce the crude death rate.</td>
</tr>
<tr>
<td></td>
<td>LTSTH</td>
<td>Selects a Coale-Demeny south model life table that will reproduce the crude death rate.</td>
</tr>
<tr>
<td></td>
<td>LTWST</td>
<td>Selects a Coale-Demeny west model life table that will reproduce the crude death rate.</td>
</tr>
</tbody>
</table>

58
<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population and deaths by 5-year age groups</td>
<td>LTOPDTH</td>
<td>Constructs a life table.</td>
</tr>
<tr>
<td></td>
<td>PRECOA</td>
<td>Evaluates the coverage of registered deaths in relation to the population (Preston-Coale method).</td>
</tr>
<tr>
<td></td>
<td>GBRAL</td>
<td>Evaluates the coverage of registered deaths in relation to the population (Brass growth balance method).</td>
</tr>
<tr>
<td>Population and death rates by 5-year age groups</td>
<td>ADJMX</td>
<td>Proportionally adjusts the age-specific death rates to obtain a desired number of total deaths.</td>
</tr>
<tr>
<td>Two census populations by 5-year age groups</td>
<td>PREBEN</td>
<td>Estimates life expectancies by age for ages 10 years and above (Preston-Bennett method).</td>
</tr>
<tr>
<td></td>
<td>PRESTO</td>
<td>U.N. MORTPAK, estimates levels of mortality and other demographic parameters (Preston integrated method).</td>
</tr>
<tr>
<td>Two census populations and intercensal deaths by 5-year age groups</td>
<td>BENHR</td>
<td>U.N. MORTPAK, estimates mortality levels and evaluates deaths (Bennett-Horiuchi method).</td>
</tr>
<tr>
<td>Average children ever born and children surviving per woman by 5-year age groups of females</td>
<td>CEBCS</td>
<td>U.N. MORTPAK, estimates infant and child mortality (modifications of Brass method).</td>
</tr>
<tr>
<td>Children born during last year and children surviving</td>
<td>BTHSRV</td>
<td>Estimates infant mortality (Johnson method).</td>
</tr>
<tr>
<td>Information</td>
<td>Spreadsheet</td>
<td>Procedure</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Age-specific death rates or probabilities of dying</td>
<td>LTMXQXAD</td>
<td>Constructs a life table from a set of ( m_x ) or ( q_x ) values.</td>
</tr>
<tr>
<td></td>
<td>LOGITQX</td>
<td>Smooths the ( q_x ) function and constructs a life table (logit method).</td>
</tr>
<tr>
<td></td>
<td>LIFTB</td>
<td>U.N. MORTPAK, constructs a life table from a set of ( q_x ) or ( m_x ) values.</td>
</tr>
<tr>
<td>Selected life table functions (( l_x ), ( q_x ), ( m_x ), or ( e_x ))</td>
<td>MATCH</td>
<td>U.N. MORTPAK, constructs a life table based on models or empirical mortality patterns.</td>
</tr>
<tr>
<td>One or more probabilities of dying by age</td>
<td>BESTFT</td>
<td>U.N. MORTPAK, estimates a complete set of probabilities of dying.</td>
</tr>
<tr>
<td>Probabilities of dying from a life table and levels of life expectancies at birth, for females</td>
<td>INTPLTF</td>
<td>Estimates female life tables for desired levels of life expectancies at birth.</td>
</tr>
<tr>
<td>Probabilities of dying from a life table and levels of life expectancies at birth, for males</td>
<td>INTPLIM</td>
<td>Estimates male life tables for desired levels of life expectancies at birth.</td>
</tr>
<tr>
<td>Survivors at exact age ( x )</td>
<td>LOGITLX</td>
<td>Smooths an ( l_x ) function and constructs a life table (logit method).</td>
</tr>
</tbody>
</table>
Spreadsheets, United Nations Programs, and Methods That Can Be Used for Analyzing Mortality, According to the Available Information--Continued

<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life expectancies at birth</td>
<td>EOLGST</td>
<td>Interpolates and extrapolates life expectancies at birth (logistic function).</td>
</tr>
<tr>
<td>Age-specific death and fertility rates</td>
<td>SP</td>
<td>Estimates a stable population and fertility rates.</td>
</tr>
</tbody>
</table>

Where to Find the Software

The following Bureau of the Census spreadsheets for measuring the level of mortality in a population can be found in volume II:

Bureau of the Census spreadsheets:

<table>
<thead>
<tr>
<th>ADJMX</th>
<th>LTMXQXAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTHSRV</td>
<td>LTNTH</td>
</tr>
<tr>
<td>GRBAL</td>
<td>LTOPDP</td>
</tr>
<tr>
<td>EOLGST</td>
<td>LTSTH</td>
</tr>
<tr>
<td>INTPLTF</td>
<td>LTWST</td>
</tr>
<tr>
<td>INTPLTM</td>
<td>PREBEN</td>
</tr>
<tr>
<td>LOGITLX</td>
<td>PRECOA</td>
</tr>
<tr>
<td>LOGITQX</td>
<td>SP</td>
</tr>
</tbody>
</table>

United Nations programs in the MORTPAK package:

BENHR MATCH
BESTFT ORPHAN
CEBCS PRESTO
LIFTB WIDOW
One of the most outstanding demographic events in this century has been the great decline in mortality. No more than 50 years ago, few persons could have predicted that populations such as those of Japan, Sweden, Switzerland, and Norway would achieve male and female life expectancy of 74 and 80 years, respectively. Although these countries have the world’s lowest mortality and their populations constitute only a small proportion of the world population, many developing countries, too, are making substantial progress in reducing mortality. In Costa Rica, Cuba, Panama, Sri Lanka, and Mauritius, for example, mortality is now as low as in some developed countries of the world. Mortality levels are becoming more independent of economic development than during the past, raising questions about what other developing countries can do to curtail their excess mortality.

One of the first steps in making plans for reducing mortality is to know not only the overall level but also the age structure of mortality and, if possible, the main causes of death. This chapter presents some of the most common techniques for measuring mortality levels. First it presents some methods based on reliable data on deaths and population, including the construction of a life table, and then it discusses techniques to evaluate data on deaths. Next it presents various techniques for indirect estimation of infant, child, and adult mortality when available data are incomplete or unreliable. Finally, it discusses the use of model life tables in the estimation of mortality levels and the use of logits in constructing life tables from incomplete data.

Direct Estimation of Mortality

When reliable information on deaths and population is available from registers and censuses, direct calculations of mortality can be made based on these data. The crude death rate is the most common and the easiest to calculate, but often more complicated measures are needed because they provide additional information. Infant mortality, in particular, is an important indicator of a country’s development. Age-specific death rates for other ages are also important in deciding which ages to target for particular programs. Life expectancy is a useful summary measure because it takes into account the mortality situation at each age yet expresses the result in a single figure. Each of these measures is discussed below.
Crude death rate (CDR)

The most common direct measure of mortality is the crude death rate, or the number of deaths per 1,000 population. It is calculated as the number of deaths occurring in a year divided by the population at midyear, times 1,000.

For example, the CDR for Chile in 1986 is obtained as follows:

\[
\left( \frac{72,209}{12,258,000} \right) \times 1,000 = 5.89
\]

There were 6 deaths per 1,000 population in Chile in 1986.

Infant mortality rate (IMR)

When vital registration data are available, the infant mortality rate is usually calculated as the ratio of the number of deaths of infants under 1 year of age to the number of live births occurring that year, times 1,000. A more refined rate would take into account a process for relating infant deaths to their actual birth cohort because in reality some of the deaths occurring each year correspond to infants born during the previous year, just as some infants born in the current year will die the following year before reaching their first birthday. (The problem of calculating infant mortality is considered below in the section on constructing a life table; see appendix III-2). For practical purposes, however, the calculation cited above is a good approximation of the IMR based on a given year's vital registration data or on an average of data for 3 consecutive years. (Other methods of estimating the infant mortality rate when complete registration data are not available are discussed in a later section.)

For example, the IMR for Chile in 1986 is obtained as follows:

\[
\left( \frac{5,220}{272,997} \right) \times 1,000 = 19.12
\]

There were 19 infant deaths per 1,000 live births in Chile in 1986.

Age-specific death rate

While mortality is very high at the early moments of life, it declines rapidly thereafter, reaching its lowest levels between 10 and 15 years of age.
In subsequent years, the older the age, the higher the mortality (see figure III-1-A). Because of the scale of variations of the rates, a regular graph does not perfectly portray the changes in the age-specific death rates from age to age. For a better portrayal, the values of the rates are plotted on a logarithmic scale, as shown in figure III-1-B.
Figure III-1. Age-Specific Death Rates as Shown on Arithmetic and Logarithmic Graphs

Arithmetic graph

Logarithmic graph
Because mortality varies with age, a comparison between countries based only on the crude death rate may be misleading. Two populations may have different crude death rates even if mortality at each age is the same in each of them. This would occur when the age structures of the two populations are different: the crude death rate would be higher in the population with a larger proportion of its people in the high-risk ages of mortality.

In another case, one population may actually have lower mortality at each age, and still have a higher crude death rate. At first it may strike one as surprising, for example, that although Switzerland has one of the lowest mortality levels in the world, its 1987 crude death rate of 9 per 1,000 is higher than that of the Philippines, at 7 per 1,000. This is because Switzerland has an "older" population, with 15 percent of its people over 65 years of age (where death rates are highest), while the Philippines has only 3 percent of its population over age 65 years. On the other hand, less than 6 percent of Switzerland's population is age 10 to 14 years (where mortality is usually at the lowest level) while 12 percent of the Philippines' population is in these low mortality ages.

Because of this unseen effect of the age structure on the crude death rate, other indices have been developed to analyze the levels of mortality in a population. Age-specific death rates provide a measure of mortality at each age.

Age-specific death rates are calculated as the number of deaths in a particular age group per 1,000 population in the same age group.

In symbols:

\[
M_x^t = \frac{D_x^t}{P_x^t} \times 1,000
\]

Where:

- \(M_x^t\) is the age-specific death rate between ages \(x\) and \(x+n\) for year \(t\);
- \(D_x^t\) is the number of deaths between ages \(x\) and \(x+n\) for year \(t\); and
- \(P_x^t\) is the population between ages \(x\) and \(x+n\) for year \(t\).
Although each age-specific death rate properly indicates the mortality level of a particular age group, it is cumbersome to deduce from them a general level of mortality for all ages. If the age-specific death rates in one population are higher at all ages than those of another population, the former may be said to have higher mortality than the latter. But if the levels of the age-specific death rates are higher at some ages and lower at others, it is less obvious which population has higher or lower mortality overall.

There are two ways to overcome this difficulty: (a) age-specific death rates may be used to "standardize" the crude death rate between two or more populations; and (b) age-specific death rates may be summarized into a single figure, such as the life expectancy at birth. Two methods of standardization are presented below, followed by a discussion of the life expectancy at birth.

**Standardization of Crude Death Rates**

The standardization process consists of performing a few calculations using age-specific death rates and a population age structure for the purpose of eliminating from the crude death rate the effect of different age structures between populations. There are two standardization procedures: direct and indirect (Spiegelman, 1968). The choice between the two depends on the amount of base information available.

**Direct standardization**

The purpose of direct standardization is to calculate a crude death rate in two or more populations by applying each one's own set of age-specific death rates to a single population age structure, called the "standard" population. The resulting standardized crude death rates allow a comparison of mortality in the areas independent of their particular age structures. Direct standardization of crude death rates is often used in a historical analysis of mortality trends in a specific population.

Direct standardization may be used when information on age-specific death rates is available for each population to be standardized. The first step is to select a population age distribution (if possible by 5-year age groups) to be used as the "standard." The standard may be the actual population of one of the areas, an average of all of them, or another population. The age-specific death rates of each area are then multiplied by the number of persons in each age group in the standard population, resulting in a set of "expected" deaths in each age group based on the mortality level of each population. The sum of the deaths are then divided by the total of the standard population to give the standardized crude death rate pertaining to the mortality of each area. Since only one population (the standard) is used, differences among the standardized crude death rates are due only to the mortality levels of each area (see table III-1).
Direct standardization of the crude death rate is accomplished by applying the age-specific death rates of each area's own population to the number of persons in each age group of the standard population.

In symbols:

\[
s_{d_i}^t = \frac{\sum_{x=0} W \cdot n_{x,i}^t \cdot n_{x,s}}{P_s}
\]

Where:

- \( s_{d_i}^t \) is the standardized crude death rate for area \( i \) and year \( t \);
- \( W \cdot n_{x,i}^t \) is the age-specific death rate for ages \( x \) to \( x+n \) for area \( i \) and year \( t \);
- \( n_{x,s} \) is the standard population at ages \( x \) to \( x+n \);
- \( P_s \) is the total standard population; and
- \( w \) represents the oldest age.

(See example in appendix III-1.)

In a comparative analysis of standardized rates, the higher the rate, the higher the mortality of the area it pertains to. Although standardized crude death rates may reveal whether the mortality level of one area is higher or lower than that of another area, the procedure has certain limitations. For example, use of a different standard may change the relative values of the standardized rates (see figure III-2). In addition, interpretation of the results is often difficult. For a sample calculation, see appendix III-1.

**Decomposition of the difference between two crude death rates**

The principle behind the standardization procedure can be used to "decompose" the difference between two crude death rates into its component parts, namely the portion of the difference due to the age structures and the portion due to actual mortality differentials (plus a small residual that is due to the interaction between mortality and age structure differentials) (Kitagawa, 1955).
Table III-1. Actual, Directly and Indirectly Standardized Crude Death Rates (per thousand)

<table>
<thead>
<tr>
<th>Item</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td>10.4</td>
<td>7.6</td>
<td>4.9</td>
<td>7.4</td>
<td>8.9</td>
<td>12.7</td>
</tr>
<tr>
<td>Direct Standardization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population A</td>
<td>10.4</td>
<td>6.3</td>
<td>4.6</td>
<td>6.3</td>
<td>3.4</td>
<td>3.0</td>
</tr>
<tr>
<td>Population F</td>
<td>24.6</td>
<td>19.8</td>
<td>16.1</td>
<td>15.9</td>
<td>13.6</td>
<td>12.7</td>
</tr>
<tr>
<td>Indirect Standardization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality F</td>
<td>44.3</td>
<td>26.2</td>
<td>18.6</td>
<td>24.3</td>
<td>13.7</td>
<td>12.7</td>
</tr>
</tbody>
</table>

Source: Appendix III-1.

Figure III-2. Actual and Directly Standardized Crude Death Rates

Rates (per thousand)

- Actual
- Standard A
- Standard F

Note: Standard A and Standard F represent standardized crude death rates based on standard population A and F, respectively.
Consider two populations, A and B, each one with information on age-specific death rates and population age structure, and calculate the decomposition as follows:

1. Calculate an ordinary crude death rate for areas A \([d(a,a)]\) and B \([d(b,b)]\) in the usual way and determine the difference between them.

2. Apply the age-specific death rates of area A to the population age structure of area B and determine the corresponding crude death rate. This is the "mixed" crude death rate; call it \(d(b,a)\).

3. Compare the "mixed" crude death rate \(d(b,a)\) with the ordinary crude death rate of area A, \(d(a,a)\); the discrepancy represents the portion of the difference between the crude death rates of A and B that is due to age structure differentials (because both rates have the mortality of A).

4. Apply the age-specific death rates of area B to the population age structure of area A and calculate the corresponding crude death rate. This is a new "mixed" crude death rate; call it \(d(a,b)\).

5. Compare the "mixed" crude death rate \(d(a,b)\) with the ordinary crude death rate of area A, \(d(a,a)\); the discrepancy represents the portion of the difference between the crude death rates of A and B that is due to mortality differentials (because both rates have the age structure of A).

The sum of the two portions of the difference as calculated in (3) and (5) will not be the same as the total difference between the crude death rates of populations A and B. The discrepancy is due to the interaction of mortality and age structure (see table III-2).
Table III-2. Decomposition of the Difference Between Two Crude Death Rates

<table>
<thead>
<tr>
<th>Age</th>
<th>Populations</th>
<th>Mortality rates</th>
<th>Deaths</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A (70,812)</td>
<td>B (4,423)</td>
<td>A (0.0875)</td>
<td>B (0.0187)</td>
<td>D(a,a)</td>
<td>D(a,b)</td>
<td>D(b,b)</td>
<td>D(b,a)</td>
</tr>
<tr>
<td>Under 1</td>
<td>248,032</td>
<td>16,874</td>
<td>0.0091</td>
<td>0.0009</td>
<td>6196</td>
<td>1324</td>
<td>83</td>
<td>387</td>
</tr>
<tr>
<td>1-4</td>
<td>259,475</td>
<td>20,305</td>
<td>0.0024</td>
<td>0.0005</td>
<td>623</td>
<td>130</td>
<td>10</td>
<td>49</td>
</tr>
<tr>
<td>5-9</td>
<td>216,609</td>
<td>25,132</td>
<td>0.0018</td>
<td>0.0004</td>
<td>390</td>
<td>87</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td>10-14</td>
<td>180,730</td>
<td>24,598</td>
<td>0.0028</td>
<td>0.0008</td>
<td>506</td>
<td>145</td>
<td>20</td>
<td>69</td>
</tr>
<tr>
<td>15-19</td>
<td>146,420</td>
<td>26,208</td>
<td>0.0038</td>
<td>0.0012</td>
<td>556</td>
<td>176</td>
<td>31</td>
<td>100</td>
</tr>
<tr>
<td>20-24</td>
<td>115,405</td>
<td>30,412</td>
<td>0.0042</td>
<td>0.0012</td>
<td>485</td>
<td>138</td>
<td>36</td>
<td>128</td>
</tr>
<tr>
<td>25-29</td>
<td>91,911</td>
<td>40,022</td>
<td>0.0049</td>
<td>0.0014</td>
<td>450</td>
<td>129</td>
<td>56</td>
<td>196</td>
</tr>
<tr>
<td>30-34</td>
<td>74,922</td>
<td>47,607</td>
<td>0.0058</td>
<td>0.0016</td>
<td>435</td>
<td>120</td>
<td>76</td>
<td>276</td>
</tr>
<tr>
<td>35-39</td>
<td>61,312</td>
<td>48,821</td>
<td>0.0072</td>
<td>0.0023</td>
<td>443</td>
<td>141</td>
<td>112</td>
<td>352</td>
</tr>
<tr>
<td>40-44</td>
<td>50,404</td>
<td>43,565</td>
<td>0.0094</td>
<td>0.0031</td>
<td>474</td>
<td>156</td>
<td>135</td>
<td>410</td>
</tr>
<tr>
<td>45-49</td>
<td>40,502</td>
<td>42,156</td>
<td>0.0132</td>
<td>0.0052</td>
<td>535</td>
<td>211</td>
<td>219</td>
<td>556</td>
</tr>
<tr>
<td>50-54</td>
<td>31,614</td>
<td>39,942</td>
<td>0.0187</td>
<td>0.0072</td>
<td>591</td>
<td>228</td>
<td>288</td>
<td>747</td>
</tr>
<tr>
<td>55-59</td>
<td>24,409</td>
<td>38,735</td>
<td>0.0280</td>
<td>0.0120</td>
<td>683</td>
<td>293</td>
<td>465</td>
<td>1,005</td>
</tr>
<tr>
<td>60-64</td>
<td>17,721</td>
<td>35,921</td>
<td>0.0420</td>
<td>0.0202</td>
<td>764</td>
<td>358</td>
<td>726</td>
<td>1,509</td>
</tr>
<tr>
<td>65-69</td>
<td>11,632</td>
<td>30,101</td>
<td>0.0653</td>
<td>0.0340</td>
<td>760</td>
<td>395</td>
<td>1,023</td>
<td>1,966</td>
</tr>
<tr>
<td>70-74</td>
<td>4,992</td>
<td>25,634</td>
<td>0.1017</td>
<td>0.0587</td>
<td>499</td>
<td>288</td>
<td>1,505</td>
<td>2,607</td>
</tr>
<tr>
<td>75-79</td>
<td>2,510</td>
<td>14,554</td>
<td>0.2080</td>
<td>0.1540</td>
<td>522</td>
<td>387</td>
<td>2,241</td>
<td>3,027</td>
</tr>
<tr>
<td>Total</td>
<td>1,649,522</td>
<td>555,010</td>
<td>17,148</td>
<td>4,928</td>
<td>7,052</td>
<td>13,660</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CDR: 10.40 12.71

Difference between ordinary crude rates = 10.40 - 12.71 = -2.31
Difference due to age structure = D(a,a) - D(b,a) = -14.22
Difference due to mortality = D(a,a) - D(a,b) = 7.41
Interaction = D(a,a) - D(a,b) - (D(a,a) - D(b,a)) - (D(a,a) - D(a,b)) = 4.50

Note: In the symbol D(a,b), the first index refers to the population (A or B) and the second to the mortality of each population (A or B).
Decomposition of the difference between the crude death rates of populations A and B is calculated as follows:

In symbols:

\[
\begin{align*}
\text{d}(a,b) &= \frac{\sum_{x=0}^{w} n_{x}^{M_{x},b} n_{x}^{P_{x},a}}{p_{a}} \quad \text{and} \quad d(b,a) = \frac{\sum_{x=0}^{w} n_{x}^{M_{x},a} n_{x}^{P_{x},b}}{p_{b}}
\end{align*}
\]

Where:

- \( d(a,b) \) and \( d(b,a) \) are special crude death rates with age structure of A (or B) and mortality of B (or A);
- \( n_{x}^{M_{x},a} \) and \( n_{x}^{M_{x},b} \) are the age-specific mortality rates of A and B, respectively, for ages \( x \) to \( x+n \);
- \( n_{x}^{P_{x},a} \) and \( n_{x}^{P_{x},b} \) are the populations of A and B, respectively, at ages \( x \) to \( x+n \);
- \( p_{a} \) and \( p_{b} \) are the total populations of A and B, respectively; and
- \( w \) represents the oldest age.

The portion of the difference between the crude death rates of populations A and B \([d(a,a) - d(b,b)]\) due to:

- Age structure = \( d(a,a) - d(b,a) \)
- Mortality = \( d(a,a) - d(a,b) \)
- Interaction = \( d(a,a) - d(b,b) - [d(a,a) - d(b,a)] - [d(a,a) - d(a,b)] \)

(See example in appendix III-1.)

**Indirect standardization**

Indirect standardization may be the method of choice when age-specific death rates are not available for each area whose crude death rate is to be standardized (Spiegelman, 1968). Only a population distribution by age and the total number of deaths are required for each area, and there must be available any "standard" population with information on the crude death rate and age-specific death rates.
Indirect standardization of the crude death rate is accomplished by applying the age-specific death rates of a standard population to the age structure of the area whose crude death rate is to be standardized, and then multiplying the standard crude death rate by a ratio of the area's own number of deaths to the standard number of deaths.

In symbols:

\[
i_{i}d_{a} = d_{s} \frac{D_{s}}{w} \sum_{x=0}^{n} m_{x,s} \frac{n_{p_{x,a}}}{x+n}
\]

Where:

- \(i_{i}d_{a}\) is the indirectly standardized crude death rate for population A;
- \(d_{s}\) is the crude death rate of the standard population;
- \(D_{s}\) is the total number of deaths in population A;
- \(m_{x,s}\) is the age-specific death rate of the standard population;
- \(n_{p_{x,a}}\) is the population of A at ages x to x+n; and
- w represents the oldest age.

(See example in appendix III-1.)

The procedure consists of several steps:

1. Apply the standard age-specific death rates to the population age structure of the area whose crude death rate is to be standardized, and take the sum of the resulting deaths by age. The result is the "standard" number of deaths.
(2) Divide the area's actual number of deaths by the standard number of deaths. The result is an "adjustment factor" representing the difference between the mortality of the population to be standardized and that of the standard population.

(3) Multiply the standard population's crude death rate by the adjustment factor. This will increase or decrease the standard crude death rate according to whether the mortality of the area is higher or lower than that of the standard.

As in the case of direct standardization, the results of indirect standardization would change if a different standard population were selected. The standardized rate computed by the indirect method provides a comparison of mortality between the population being studied and the standard population. However, when it is calculated for various populations using the same standard, it does not provide a direct comparison of mortality among these populations because each standardized rate depends on the age composition of the population in question. At best, mortality of the various populations may be ranked according to their relative standing in comparison to mortality of the standard population. Table III-1 and appendix III-1 illustrate both direct and indirect standardization in two populations.

Life Expectancy at Birth

One of the most useful summary measures of the overall level of mortality of a population is the life expectancy at birth. It is a more accurate reflection of mortality than the crude death rate because it is independent of the population's age structure, and it is not influenced by extraneous factors such as the selection of a standard population.

Calculation of the life expectancy at birth begins with a set of age-specific death rates, from which probabilities of surviving from one age to the next can be estimated. These survival probabilities are applied to an assumed cohort of births that occurred in the same year, following the survivors as they reach successive ages until all have eventually died. As a result of this procedure, a count can be obtained of the total number of years that the birth cohort as a whole would live under the observed mortality conditions. The ratio of all years lived by the total number of people in the cohort to the original number of births represents the average number of years to be lived by persons born in the same year under the particular mortality conditions of that year. This ratio is the life expectancy at birth.

The steps outlined above to calculate the life expectancy at birth are the steps required in the construction of a life table.
Table III-3. Life Table

<table>
<thead>
<tr>
<th>Age</th>
<th>Probability of dying</th>
<th>Survivors at age x</th>
<th>Deaths between age x and x+n</th>
<th>Survivors between age x and x+n</th>
<th>Central death rate</th>
<th>Population at age x and above</th>
<th>Survival ratio from ages x,x+n to ages x+n,x+2n</th>
<th>Life expectancy at age x</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>nq(x)</td>
<td>l(x)</td>
<td>ndx</td>
<td>nLx</td>
<td>nm(x)</td>
<td>Tx</td>
<td>nPx</td>
</tr>
<tr>
<td>0 1</td>
<td>.09371</td>
<td>100,000</td>
<td>9,371</td>
<td>91,819</td>
<td>.10206</td>
<td>5,043,286</td>
<td>.90395</td>
<td>50.43</td>
</tr>
<tr>
<td>1 4</td>
<td>.01140</td>
<td>90,629</td>
<td>1,033</td>
<td>360,154</td>
<td>.00827</td>
<td>4,951,467</td>
<td>.98634</td>
<td>54.63</td>
</tr>
<tr>
<td>5 5</td>
<td>.00972</td>
<td>89,596</td>
<td>871</td>
<td>445,800</td>
<td>.00195</td>
<td>4,591,313</td>
<td>.99093</td>
<td>51.24</td>
</tr>
<tr>
<td>10 5</td>
<td>.00842</td>
<td>88,724</td>
<td>747</td>
<td>441,754</td>
<td>.00169</td>
<td>4,145,513</td>
<td>.98812</td>
<td>46.72</td>
</tr>
<tr>
<td>15 5</td>
<td>.01538</td>
<td>87,977</td>
<td>1,353</td>
<td>636,504</td>
<td>.00310</td>
<td>3,703,758</td>
<td>.98000</td>
<td>42.10</td>
</tr>
<tr>
<td>20 5</td>
<td>.02287</td>
<td>86,624</td>
<td>1,981</td>
<td>428,169</td>
<td>.00643</td>
<td>3,267,255</td>
<td>.97324</td>
<td>37.72</td>
</tr>
<tr>
<td>25 5</td>
<td>.03074</td>
<td>84,643</td>
<td>2,602</td>
<td>416,710</td>
<td>.00624</td>
<td>2,839,086</td>
<td>.96628</td>
<td>33.54</td>
</tr>
<tr>
<td>30 5</td>
<td>.03680</td>
<td>82,041</td>
<td>3,019</td>
<td>402,657</td>
<td>.00750</td>
<td>2,422,376</td>
<td>.95673</td>
<td>29.53</td>
</tr>
<tr>
<td>5 5</td>
<td>.01140</td>
<td>90,629</td>
<td>1,033</td>
<td>360,154</td>
<td>.00827</td>
<td>4,951,467</td>
<td>.98634</td>
<td>54.63</td>
</tr>
<tr>
<td>1 4</td>
<td>.01140</td>
<td>90,629</td>
<td>1,033</td>
<td>360,154</td>
<td>.00827</td>
<td>4,951,467</td>
<td>.98634</td>
<td>54.63</td>
</tr>
<tr>
<td>5 5</td>
<td>.00972</td>
<td>89,596</td>
<td>871</td>
<td>445,800</td>
<td>.00195</td>
<td>4,591,313</td>
<td>.99093</td>
<td>51.24</td>
</tr>
<tr>
<td>10 5</td>
<td>.00842</td>
<td>88,724</td>
<td>747</td>
<td>441,754</td>
<td>.00169</td>
<td>4,145,513</td>
<td>.98812</td>
<td>46.72</td>
</tr>
<tr>
<td>15 5</td>
<td>.01538</td>
<td>87,977</td>
<td>1,353</td>
<td>636,504</td>
<td>.00310</td>
<td>3,703,758</td>
<td>.98000</td>
<td>42.10</td>
</tr>
<tr>
<td>20 5</td>
<td>.02287</td>
<td>86,624</td>
<td>1,981</td>
<td>428,169</td>
<td>.00643</td>
<td>3,267,255</td>
<td>.97324</td>
<td>37.72</td>
</tr>
<tr>
<td>25 5</td>
<td>.03074</td>
<td>84,643</td>
<td>2,602</td>
<td>416,710</td>
<td>.00624</td>
<td>2,839,086</td>
<td>.96628</td>
<td>33.54</td>
</tr>
<tr>
<td>30 5</td>
<td>.03680</td>
<td>82,041</td>
<td>3,019</td>
<td>402,657</td>
<td>.00750</td>
<td>2,422,376</td>
<td>.95673</td>
<td>29.53</td>
</tr>
<tr>
<td>5 5</td>
<td>.01140</td>
<td>90,629</td>
<td>1,033</td>
<td>360,154</td>
<td>.00827</td>
<td>4,951,467</td>
<td>.98634</td>
<td>54.63</td>
</tr>
<tr>
<td>1 4</td>
<td>.01140</td>
<td>90,629</td>
<td>1,033</td>
<td>360,154</td>
<td>.00827</td>
<td>4,951,467</td>
<td>.98634</td>
<td>54.63</td>
</tr>
<tr>
<td>5 5</td>
<td>.00972</td>
<td>89,596</td>
<td>871</td>
<td>445,800</td>
<td>.00195</td>
<td>4,591,313</td>
<td>.99093</td>
<td>51.24</td>
</tr>
<tr>
<td>10 5</td>
<td>.00842</td>
<td>88,724</td>
<td>747</td>
<td>441,754</td>
<td>.00169</td>
<td>4,145,513</td>
<td>.98812</td>
<td>46.72</td>
</tr>
<tr>
<td>15 5</td>
<td>.01538</td>
<td>87,977</td>
<td>1,353</td>
<td>636,504</td>
<td>.00310</td>
<td>3,703,758</td>
<td>.98000</td>
<td>42.10</td>
</tr>
<tr>
<td>20 5</td>
<td>.02287</td>
<td>86,624</td>
<td>1,981</td>
<td>428,169</td>
<td>.00643</td>
<td>3,267,255</td>
<td>.97324</td>
<td>37.72</td>
</tr>
<tr>
<td>25 5</td>
<td>.03074</td>
<td>84,643</td>
<td>2,602</td>
<td>416,710</td>
<td>.00624</td>
<td>2,839,086</td>
<td>.96628</td>
<td>33.54</td>
</tr>
<tr>
<td>30 5</td>
<td>.03680</td>
<td>82,041</td>
<td>3,019</td>
<td>402,657</td>
<td>.00750</td>
<td>2,422,376</td>
<td>.95673</td>
<td>29.53</td>
</tr>
<tr>
<td>5 5</td>
<td>.01140</td>
<td>90,629</td>
<td>1,033</td>
<td>360,154</td>
<td>.00827</td>
<td>4,951,467</td>
<td>.98634</td>
<td>54.63</td>
</tr>
<tr>
<td>1 4</td>
<td>.01140</td>
<td>90,629</td>
<td>1,033</td>
<td>360,154</td>
<td>.00827</td>
<td>4,951,467</td>
<td>.98634</td>
<td>54.63</td>
</tr>
<tr>
<td>5 5</td>
<td>.00972</td>
<td>89,596</td>
<td>871</td>
<td>445,800</td>
<td>.00195</td>
<td>4,591,313</td>
<td>.99093</td>
<td>51.24</td>
</tr>
<tr>
<td>10 5</td>
<td>.00842</td>
<td>88,724</td>
<td>747</td>
<td>441,754</td>
<td>.00169</td>
<td>4,145,513</td>
<td>.98812</td>
<td>46.72</td>
</tr>
<tr>
<td>15 5</td>
<td>.01538</td>
<td>87,977</td>
<td>1,353</td>
<td>636,504</td>
<td>.00310</td>
<td>3,703,758</td>
<td>.98000</td>
<td>42.10</td>
</tr>
<tr>
<td>20 5</td>
<td>.02287</td>
<td>86,624</td>
<td>1,981</td>
<td>428,169</td>
<td>.00643</td>
<td>3,267,255</td>
<td>.97324</td>
<td>37.72</td>
</tr>
<tr>
<td>25 5</td>
<td>.03074</td>
<td>84,643</td>
<td>2,602</td>
<td>416,710</td>
<td>.00624</td>
<td>2,839,086</td>
<td>.96628</td>
<td>33.54</td>
</tr>
<tr>
<td>30 5</td>
<td>.03680</td>
<td>82,041</td>
<td>3,019</td>
<td>402,657</td>
<td>.00750</td>
<td>2,422,376</td>
<td>.95673</td>
<td>29.53</td>
</tr>
</tbody>
</table>

Note: The first nPx is the survival ratio from birth to ages 0-4, the second from ages 0-4 to 5-9, and the last one to the open-ended age group 80+. 
Life Tables

A life table serves useful purposes both within the demographic community and in the world at large. As noted above, it is the source of estimates of the life expectancy at birth. In addition, it provides survival ratios for each age or age group that are used in making population projections. Life insurance companies use life tables (which they may call actuarial tables) to determine their clients' probable life spans and hence their insurance premiums according to their particular characteristics.

The construction of a life table may at first seem to be a formidable task, but in fact the steps are quite routine. Some detailed aspects of the calculations are presented in appendix III-2, while the basic life table concepts and functions are presented in this section.

A life table follows a hypothetical cohort of 100,000 persons born at the same time (called the "radix" of the life table) as they progress through successive ages, with the cohort reduced from one age to the next according to a set of actual death rates by age, until all persons eventually die. A complete, or unabridged, life table is constructed by single years of age, while an abridged life table is constructed by 5-year age groups. A life table can be constructed for both sexes together or, more commonly, for each sex separately.

The construction of a life table consists of calculating various interdependent "functions," using as a base the available age-specific death rates. (The base data, both the reported deaths and the population figures on which the death rates are based, should be evaluated before the life table construction begins. The next section offers some guidelines.)

The first function calculated in the life table is the probability of dying between two exact ages, for example, the probability that a person of exact age 30 will die before reaching his or her 35th birthday. This function is symbolized as \( n_{qx} \), where \( x \) represents exact age \( x \) and \( n \) is the age interval. In the above example, the symbol would be \( 5_{30} \).

The probability of dying is obtained by using age-specific death rates and "separation factors" of death. The separation factors represent the average number of years lived during any age interval by persons who die between the defined limits of the age interval. For example, persons who die at age 23 (between exact ages 23 and 24) would live about half that year, on average (since deaths are fairly evenly distributed throughout the year); thus the separation factor in this case would be close to 0.5. Similarly, when data are presented in 5-year age groups, persons who die between ages 35 and 40 would live on average about 2.5 years during the age interval, and that would be the corresponding separation factor. (Further details on the calculation of separation factors are presented in appendix III-2.)
In applied demography, the only $q_0$ function that is frequently used is $q_0$, representing the probability of dying between birth and age 1; this is usually called the infant mortality rate. Other $q_x$ functions are more pertinent to actuarial analysis or demographic theory, but they are all required for calculating the rest of the life table functions. (See figure III-3 and table III-3.)

Recently, the use of $q_0$ has increased in developing countries as an index for obtaining a better trend of child mortality. A technique for an indirect estimation of child mortality is presented later.

Probability of dying between exact ages: $q_x$

In symbols:

$$ q_x = - \frac{n \cdot M_x}{l + (n - k_x) \cdot M_x} $$

Where:

$nM_x$ is the age-specific death rate for ages $x$ to $x+n$; and

$nk_x$ is the separation factor of deaths for ages $x$ to $x+n$.

The next life table function to be calculated is the number of persons surviving to each exact age. As the life table usually starts with a radix of 100,000 births (at exact age 0), the number of survivors from birth to each exact age is obtained using the probabilities as estimated above. The number of survivors at each exact age is represented by the symbol $l_x$ (see figure III-3).
Figure III-3. Selected Life Table Functions

A. 10^x Survivors at exact age x

B. Survival ratios \( nP_x \) and Probabilities of dying \( nq_x \)
Survivors at exact age \( x \): \( l_x \)

In symbols:
\[
l_{x+n} = l_x \left(1 - nq_x\right)
\]

Where:
- \( nq_x \) is as defined above; and
- The first \( l_x \) is \( l_0 \) and usually is defined as 100,000.

Since \( l_x \) represents the number of persons alive at each exact age \( x \), the difference between two consecutive values (for example, \( l_x \) and \( l_{x+n} \)) represents the number of deaths between the corresponding ages (\( x \) and \( x+n \), in this case). This number of deaths between two exact ages is symbolized by \( n d_x \) in the life table.

Deaths between two exact ages: \( n d_x \)

In symbols:
\[
n d_x = l_x - l_{x+n}
\]

Where:
- \( l_x \) is as defined above.

Actual populations are usually enumerated in a census or survey as the number of persons alive between two ages, for example, the population at age 23 (persons between their 23rd and 24th birthdays) or the population in the age group 15 to 19 years (persons who have reached their 15th birthday but not yet their 20th). A life table includes an analogous population, represented by the \( n l_x \) function, referred to as the "life table population" between exact ages \( x \) and \( x+n \), or the number of survivors between exact age \( x \) and \( x+n \). For example, the symbol \( s L_{55} \) refers to the life table population at ages 55 to 59 years (persons who have reached their 55th birthday but not yet their 60th).
This life table population may be interpreted in three different ways, each related to the basic concept of the life table itself: following through time a cohort of births under mortality conditions observed during a year. The three interpretations are as follows:

(1) The life table population may be interpreted as a "stationary" population. This means that every year there are 100,000 births and the same number of deaths, mortality and fertility remain constant, and the number of people in each age group \( nL_x \) does not change.

(2) Persons in each particular age group, say 30 to 35 years \( (5 L 30) \) may be interpreted as the survivors of 500,000 births that occurred during a 5-year period (100,000 each year) 30 to 35 years ago.

(3) Finally, the total life table age distribution may be interpreted as the survivorship pattern of a single cohort of 100,000 births as it passes through all ages. In this case, the value of \( nL_x \) in any particular age group represents the number of years that the single cohort of 100,000 would live between two specific ages, \( x \) to \( x+n \). For example, the \( 5L_{30} \) value refers to the total number of "person-years" that 100,000 births would live between exact ages 30 and 35.

The \( nL_x \) function is calculated as the average of two consecutive \( l_x \) values that represent persons alive at two specified exact ages.

\[
\text{Life table population: } nL_x
\]

In symbols:

\[
nL_x = n_kx \ l_x + (n - n_kx) \ l_{x+n}
\]

Where:

\( n_kx \) and \( l_x \) are as defined above.

The next function to be calculated refers to the age-specific death rates of the life table population, which are also called central death rates. These rates are derived by dividing the life table number of deaths between two specific ages by the life table population between the same ages.
Central death rates: \( n^m_x \)

In symbols:

\[
\frac{n^d_x}{n^L_x}
\]

Where:

\( n^d_x \) and \( n^L_x \) are as defined above.

Next, the "total population" of the life table may be calculated by summing all the \( n^L_x \) values. As mentioned earlier, when the life table is conceived as the following of a single cohort through time, the \( n^L_x \) values represent the total number of years to be lived by the original birth cohort of 100,000 until all persons have eventually died. A cumulative summation of the \( n^L_x \) values from the oldest ages to the youngest one represents, at any given age \( x \), the total number of years remaining to be lived by all persons who are still alive at age \( x \); it also represents the population age \( x \) years and over. The cumulative values for each age are represented by the symbol \( T_x \).

Person years of life remaining for ages \( x \) and above: \( T_x \)

In symbols:

\[
T_x = \sum_{i=x}^{w} n^L_i
\]

Where:

\( n^L_x \) is as defined above; and

\( w \) represents the oldest age.

In the case of an abridged life table, the \( n^L_x \) value for the oldest age group \( (L_w) \) refers to an open-ended age (85 years and over, for example, as in table III-3) and requires selection of a special estimating procedure; this is discussed in appendix III-2.
Now that the most commonly used life table functions have been calculated, the life expectancy can be derived. The ratio of the number of years that the life table population will live from age \( x \) up to the point when all have died, to the number of persons alive at exact age \( x \), represents the average number of years remaining to be lived by those who are alive at each age \( x \). This is the life expectancy at any given age, symbolized by \( e_x \). For example, \( e_{35} \) represents the number of years of life remaining for a person age 35, while \( e_0 \) represents the life expectancy at birth.

Life expectancy at age \( x \): \( e_x \)

In symbols:

\[
e_x = \frac{T_x}{l_x}
\]

Where \( T_x \) and \( l_x \) are as defined above.

Finally, there is an optional life table function that is used in making population projections, known as the survival ratio. Since the \( 5L_x \) function is the number of persons alive between ages \( x \) and \( x+5 \), the ratio of two consecutive \( 5L_x \) values represents the survival between the two age groups. For example, at the end of a 5-year period, persons in the group \( 5L_{25} \) are the survivors of the group \( 5L_{20} \), and hence the ratio of \( 5L_{30} \) divided by \( 5L_{25} \) represents the average chance that a person in the age group 25 to 29 has of surviving 5 years to ages 30 to 34. This ratio is symbolized by \( 5P_{25} \) and is called the 5-year survival ratio for ages 25 to 29 years (see figure III-3).

Survival ratio: \( 5P_x \)

In symbols:

\[
5P_x = \frac{5L_{x+5}}{5L_x}
\]

Where:

\( 5L_x \) is as defined above.
As noted above, this ratio is used for projecting populations. For example, if the number of persons ages 25 to 29 in the year 1980 was 3,758 and the survival ratio for 1980 for ages 25 to 29 was 0.98, the product of the population times the survival ratio, 3,683, would represent the surviving population in 1985 at ages 30 to 34 years.

The youngest and oldest age groups require special consideration. The first survival ratio represents the survival of births to ages 0 to 4 years.

\[
\text{Survival ratio from birth (b) to ages 0 to 4: } sP_b
\]

\[
sP_b = \frac{sL_0}{5 \times l_0}
\]

Where:

\(sL_0\) and \(l_0\) are as defined above.

The oldest, or open-ended, age group is represented by \(P_w\).

\[
\text{Survival ratio of open-ended age group: } P_w
\]

\[
P_{w-5} = \frac{T_w}{T_{w-5}}
\]

Where:

\(T\) is as defined above; and

\(w\) represents the oldest age.

There are several programs that perform all the above calculations for constructing a life table. Differences among these programs are due to differences in the available information. (See Bureau of the Census spreadsheets LTP0PDTH and LTMXQXAD in volume II, and the United Nations program LIFTB in the MORTPAK package.)
Evaluation and Adjustment of Death Rates

As mentioned elsewhere, age-specific death rates may be calculated when registered data on deaths and on the corresponding population by age and sex are available. Even when such data appear to be reliable, however, they should not be accepted blindly when being used for purposes such as constructing a life table. Chapter II discussed various possibilities of age misreporting in population data; such errors may be even more prevalent in age reporting of deaths, causing irregularities in the pattern of age-specific death rates that do not correspond to reality.

Detecting errors in age-specific death rates

Except for peculiar transitory conditions, death rates by age in most populations follow a rather standard pattern, and so a graph of the calculated rates usually provides a means for detecting possible errors in age misreporting. For example, table III-4 presents age-specific death rates calculated for a country whose information on population and deaths is accepted as complete. The rates in column 3 of the table seem to follow the expected mortality pattern by age: starting with relatively high mortality for infants, the rates decline to a minimum value in the age group 10 to 14 years and then increase with age. However, a semilogarithmic graph (figure III-4-A) of these rates reveals certain irregularities in the adult ages. Since the increasing mortality by age is not expected to have such a zigzag pattern, the rates need some smoothing before being used to construct a life table.

Smoothing death rates

Death rates may be smoothed in various ways, but simple techniques are usually the best. Even a free-hand graph can be used, but it is subject to different interpretations by different analysts and so the procedure can be risky.

A moving average formula usually works well. As it will be applied in an instance where overall death registration is judged to be complete, the result should retain the original registered total number of deaths pertaining to the span of age groups being smoothed. After the young adult ages (15 or 20 years), the pattern of death rates by age follows an approximately exponential shape (see figures III-1-A and III-1-B). Because of this exponential shape, a simple arithmetic average for smoothing would produce rates that overestimate the level of mortality. Therefore it is recommended to take logarithms of the death rates before calculating the average. For example, smoothing rates for ages 20 and over presented in figure III-4-A would require the following steps:

1. Take logarithms of the rates in the proper ages.
2. Average each consecutive three logarithms of the rates.
Table III-4. Reported and Smoothed Age-Specific Death Rates

<table>
<thead>
<tr>
<th>Age</th>
<th>Population</th>
<th>Deaths</th>
<th>Death rates</th>
<th>Smoothed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Total</td>
<td>18,268,160</td>
<td>445,571</td>
<td>.1021</td>
<td>.1021</td>
</tr>
<tr>
<td>Under 1</td>
<td>336,902</td>
<td>34,384</td>
<td>.0029</td>
<td>.0029</td>
</tr>
<tr>
<td>1-4</td>
<td>1,406,121</td>
<td>4,035</td>
<td>.0020</td>
<td>.0020</td>
</tr>
<tr>
<td>5-9</td>
<td>1,665,345</td>
<td>3,254</td>
<td>.0017</td>
<td>.0017</td>
</tr>
<tr>
<td>10-14</td>
<td>1,511,178</td>
<td>2,556</td>
<td>.0035</td>
<td>.0035</td>
</tr>
<tr>
<td>15-19</td>
<td>1,292,514</td>
<td>4,475</td>
<td>.0049</td>
<td>.0049</td>
</tr>
<tr>
<td>20-24</td>
<td>1,343,325</td>
<td>6,743</td>
<td>.0056</td>
<td>.0056</td>
</tr>
<tr>
<td>25-29</td>
<td>1,638,580</td>
<td>9,223</td>
<td>.0085</td>
<td>.0085</td>
</tr>
<tr>
<td>30-34</td>
<td>1,690,564</td>
<td>14,387</td>
<td>.0103</td>
<td>.0103</td>
</tr>
<tr>
<td>35-39</td>
<td>1,471,949</td>
<td>12,598</td>
<td>.0142</td>
<td>.0142</td>
</tr>
<tr>
<td>40-45</td>
<td>975,943</td>
<td>13,265</td>
<td>.0226</td>
<td>.0226</td>
</tr>
<tr>
<td>45-49</td>
<td>994,403</td>
<td>22,098</td>
<td>.0322</td>
<td>.0322</td>
</tr>
<tr>
<td>50-54</td>
<td>941,078</td>
<td>40,740</td>
<td>.0410</td>
<td>.0410</td>
</tr>
<tr>
<td>55-59</td>
<td>736,699</td>
<td>32,847</td>
<td>.0549</td>
<td>.0549</td>
</tr>
<tr>
<td>60-64</td>
<td>438,006</td>
<td>40,746</td>
<td>.0735</td>
<td>.0735</td>
</tr>
<tr>
<td>65-69</td>
<td>386,767</td>
<td>53,810</td>
<td>.0992</td>
<td>.0992</td>
</tr>
<tr>
<td>70-74</td>
<td>300,042</td>
<td>44,879</td>
<td>.1391</td>
<td>.1391</td>
</tr>
<tr>
<td>75-79</td>
<td>142,667</td>
<td>33,756</td>
<td>.2366</td>
<td>.2366</td>
</tr>
<tr>
<td>80-84</td>
<td>54,987</td>
<td>17,653</td>
<td>.3210</td>
<td>.3210</td>
</tr>
<tr>
<td>85+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

85
Figure III-4. Unsmoothed and Smoothed Age-Specific Death Rates

Unsmoothed rates

Smoothed rates

Age
(3) Take antilogarithms of the averages.

(4) Multiply the rates derived in (3) by the corresponding populations to calculate the number of deaths at each age.

(5) Sum the calculated deaths and compare them to the total registered deaths for ages 20 to 74.

(6) Take a ratio of registered to calculated deaths as compared in (5).

(7) Adjust the rates of (3) by multiplying them by the ratio calculated in (6).

The smoothed and adjusted age-specific deaths rates may then be used to construct a life table. See figure III-4-B for the smoothed rates and Bureau of the Census spreadsheet LTPOPDTH (volume II) to make these calculations.

Other Possibilities for Selecting or Calculating a Life Table

In some cases, limited data can be used to construct a life table or to estimate a life expectancy at birth to represent the mortality conditions of a population (mainly populations of small areas of a country).

For example, a country may have a set of age-specific death rates for the whole country or for a particular state, but only total deaths (not by age) for subareas, such as municipios or counties. If the population of the subarea is available by age, and if it is assumed that the pattern of mortality for the subarea is similar to the available pattern of mortality, a life table can be estimated for the subarea. This life table is calculated by adjusting the available set of death rates in such a way that, when the rates are multiplied by the population of the subarea, the sum of the products reproduces the total number of deaths of the subarea (see appendix III-3). This procedure is performed by the microcomputer spreadsheet ADJMX (see volume II).

If a pattern of mortality (as represented by a set of age-specific death rates) is not available, any other pattern could be used as long as it can be taken to represent the pattern of mortality of the subarea's population. Consequently, in some cases, model patterns can be used in the procedure described above. Three microcomputer spreadsheets use Coale-Demeny model life tables pertaining to regions North, South, and West, to make such calculations. They are LTNTH, LTSTH, and LTWST, respectively, and require the male and female population by 5-year age groups and a crude death rate for each sex or both sexes combined (see appendix III-3 and documentation of spreadsheets in volume II).
After selecting a model life table, these spreadsheets also calculate other indices, such as the crude birth rate and the infant mortality rate. The estimate of the crude birth rate is sensitive to the completeness of enumeration of the population under 1 year of age, and it is provided mainly as a control for the information used. For instance, in most developing countries, the population under 1 year of age is underreported, and the birth rate calculated by these spreadsheets will be lower than expected. Hence, before accepting an estimate of the life expectancy at birth and the model life table given by the spreadsheets, it is advisable to increase the population under 1 year of age to obtain an acceptable level of the birth rate. The correction of the infant population does not have a large impact on the estimation of the life expectancy at birth and model life table.

However, if the underenumeration of the population under 1 year of age is high, the crude death rate may not take into account all infant deaths. Unfortunately, each case is different and no generalization can be made to take this aspect into account.

Increasing the population under 1 year of age without modifying the crude death rate will increase the life expectancy at birth. In contrast, an increase in the crude death rate will decrease the life expectancy at birth. If both the population under 1 year of age and the crude death rate are underreported, part of the impact of such omissions will cancel out. The different effects of such underreporting of population and crude death rate can be analyzed by changing the population and the crude birth rate in the spreadsheet.

Another possibility exists for adjusting a pattern of mortality to obtain a more representative pattern for a particular area or country. Death registration in developing countries is usually incomplete, and underregistration of deaths is not the same in all ages. Frequently, underregistration of infant deaths and deaths among children in very young ages is higher than for other ages. If a pattern of mortality is available, and if the degree of completeness of death registration by age is known, then the mortality pattern can be adjusted accordingly. The mortality pattern can be represented by a set of age-specific death rates or by probabilities of dying. The spreadsheet LTMXQXAD performs such adjustments and calculates a life table with the adjusted pattern of mortality (see appendix III-2 and documentation in volume II).

Techniques for evaluating and adjusting data on deaths

The previous section considered various indices for analyzing mortality under the condition that information on mortality and population is reliable. Unfortunately, many populations, particularly in developing countries, still do not have a vital registration system that provides information of the required quality or completeness for calculating reliable demographic estimates. In some cases, however, an evaluation of the existing data will suggest that they are adequate if certain adjustments are made. Techniques for evaluating and adjusting such data are presented in this section. In other cases, no vital registration data exist at all, or the available data
are too sparse even for adjustment. Under such conditions, mortality must be estimated indirectly. Indirect estimation is the subject of the next section.

Various techniques have been developed to evaluate and correct information on deaths by sex and age in relation to information on population. Data on deaths may be provided not only in vital statistics registers, but also in surveys or censuses that include questions concerning deaths during a specific period of time, for example, deaths of any household members during the past year. If registered deaths can be evaluated and adjusted for errors, they can be used to obtain valuable information about the level and pattern of mortality.

**Stable population theory**

A number of the techniques for evaluating and adjusting data on mortality are based on the assumption that the population is "stable." A stable population is one in which there has been no migration, and neither fertility nor mortality has changed in the past (Dublin and Lotka, 1925; Lotka, 1934, 1937, and 1939; and United Nations, 1968).

A "stationary" population is a special kind of stable population in which not only have fertility and mortality not changed in the past but crude birth and death rates are equal to one another. Remember from the discussion of the \( nL_x \) life table function that the life table population represents a stationary population in which fertility and mortality do not change and in which the crude birth rate is the same as the crude death rate. As a result of these characteristics, combined with the absence of migration, the stationary population does not grow, and the age distribution does not change over time. Under these conditions, the number of persons alive at age \( x \) is equal to the annual number of deaths at ages \( x \) and above.

If the population whose completeness of death registration is to be evaluated were a stationary population, the evaluation procedure would be a simple one. If deaths were underregistered in such a population, then the sum of the annual number of registered deaths of age \( x \) and above would be smaller than the population at age \( x \); and the ratio of the registered deaths to the population would represent the completeness of death registration. Although this is a useful concept to consider at the outset, real populations are not stationary and so alternatives must be sought.

Since the birth and death rates of a stable population are not necessarily equal (that is, a stable population does not have to be stationary), a stable population might be growing. Although the birth and death rates do not change over time, the two rates are not necessarily at the same level. In this sense, at least some real populations are close to being stable and thus correspond to the situation implicit in some of the evaluation techniques. (For calculating a stable population, see appendix III-4 and Bureau of the Census software program SP; see documentation in volume II.)
**Preston-Coale technique**

One of the techniques to estimate the completeness of registered deaths in relation to population data was developed by Preston and Coale. It requires that information be available on the population growth rate and on both deaths and population by 5-year age groups. It assumes that the population is stable, in other words, that mortality and fertility have not changed during the past and that there has been no migration.

In a stable population with a positive growth rate, the population at age x is greater than the sum of annual deaths for ages x and above. In other words, since the population is growing, the annual deaths over age x pertain to a smaller cohort of births than the cohort of births from which the population at age x has survived. Thus, if the actual population is assumed to be stable, an adjustment of registered deaths is needed so that the sum of deaths for age x and above can be compared to the population age x. Based on stable population theory, the adjustment of deaths is made by multiplying deaths by an exponential factor (derived using the population's growth rate and the mean age of the age group). Then, the deaths are cumulated and taken as an estimate of the population at a certain age. This estimate is compared with the actual population at the same age. The ratio of the estimated to the actual population represents the completeness of registered deaths (United Nations, 1983).

While the technique is useful in particular situations, the implicit assumptions about constant mortality and fertility, as well as the absence of migration, may depart considerably from actual conditions in many populations. In practically all countries, mortality has begun to decline and in most of them fertility has also started to change. International migration has often become important as well. Furthermore, the technique requires a knowledge of the growth rate, which has an important role and impact in estimating the population from registered deaths. Finally, results of the technique will be biased if age misreporting of deaths is different from age misreporting of population (see further aspects of this technique in appendix III-5). The calculation of this technique can be performed by using the Bureau of the Census spreadsheet PRECOA; see documentation in volume II.

**Growth balance technique**

Another technique for evaluating the completeness of death registration in relation to population data was developed by Brass. Like the Preston-Coale technique, it is based on stable population theory and thus has the same assumptions of constant fertility and mortality and the absence of migration. However, the growth balance equation has an advantage over the previous technique in that it does not require knowledge of the population growth rate. In fact, if the population meets the assumptions and in addition has no age misreporting, the technique provides an estimate not only of the completeness of death registration but of the population growth rate as well (Brass, 1975; United Nations, 1983).
This technique is based on the basic observation that the birth rate of a population equals the growth rate plus the death rate; hence the name: "growth balance equation" method. For each group age x years and over (for example, 5 years and over, 10 years and over, 15 years and over), the technique estimates: (a) a "birth rate" calculated as the ratio of persons at exact age x to the population age x and over; and (b) a "death rate" calculated as the ratio of deaths ages x and over to the population of the same ages. If the population is stable, the birth and death rates for each cumulative age group are linearly related (forming a straight line on a graph). The y-intercept of this line is the population growth rate, and the slope (the coefficient of the death rate) represents the reciprocal of the completeness of death registration, or an adjustment factor for the number of deaths. In a particular analysis, if all the points actually follow a linear trend, the results may be accepted.

Real populations, of course, often diverge from the ideal conditions for applying the technique. Populations usually are not precisely stable, there is often age misreporting of the population and of deaths, and there is often differential completeness in the registration of population and of deaths by age. As a result, the pairs of birth and death rates for each age group may not fall in a straight line (see figure III-5). In such a case, the author of the technique suggests separating the points into two groups, computing the average birth rate and death rate in each, and then fitting a straight line to the pairs of points. The slope of the adjusted line would thus represent an average adjustment factor for registered deaths.

When this adjustment is required, a decision must be made as to which points should be used for fitting the straight line to the data to determine the final adjustment factor, that is, whether to use all the points or only those resembling a straight line. For example, if the population's fertility has started to decline, points calculated using very young ages may be biased due to this violation of the assumptions. Under this circumstance, it would be desirable to eliminate such points from the analysis. However, there are no rigid criteria for determining which points should be omitted and which ones included, and the estimated completeness of death registration could vary significantly depending on the points selected for fitting the straight line (see figure III-6, A and B). In general, the greater the departure from linearity, the greater the caution in accepting the results (see appendix III-6). The Bureau of the Census developed the spreadsheet GRBAL for applying this technique; see documentation in volume II.

**Bennett-Horiuchi technique**

A more recent technique by Bennett and Horiuchi (1981) for estimating underregistration of deaths does not assume that the population is stable, but more information is required for its application: a distribution of population by 5-year age groups from two censuses and registered deaths during the intercensal period. It assumes that completeness and age misreporting are the same in the two censuses and that migration is nil. However, if the population has been exposed to migration during the intercensal period, the technique can still provide acceptable results if the age and sex
characteristics of the migrants are known so migrants can be subtracted from the population figures.

This technique estimates a population for each age group and compares it with the enumerated population in the same age group. The estimated population is based on registered deaths, a population at each exact age derived from census information, and an intercensal growth rate of the population in each age group.

Like the other techniques, this one is affected by migration. Immigration will cause death registration to appear to be more complete, while emigration will cause it to appear less complete. Also, differential coverage of the population between the two censuses will have an impact. The population of both censuses is used to calculate growth rates, which in turn are used together with data on deaths to obtain an estimated population to be compared with the census population. If the first census is underenumerated in relation to the second, death registration will appear to be more complete; conversely, if the second census is underenumerated in relation to the first, death registration will appear to be less complete. Finally, the technique assumes no age misreporting under age 50 years; thus, if these ages were not reported properly, the results may be biased. This technique is presented in appendix III-7, and a computer program that may be used to apply this technique is called BENHR in the United Nations MORTPAK package.

Figure III-5. Growth Balance Equation: Burkina Faso, 1976

![Graph showing the growth balance equation for Burkina Faso, 1976. Numbers inside the figure represent ages.](image)

Note: Numbers inside figure represent ages.
Figure III-6. Age-Specific Death Rates According to the Growth Balance Method

\[ \frac{N(x)}{N(x+)} \]

Adjustment based on ages 5 to 60

- Actual
- Adjusted

Note: Numbers inside figures represent ages.

93
In addition to the techniques presented in this section, other related techniques are available, modifying slightly the assumptions of the first two techniques presented here. Although the results are similar, the assumptions as well as errors in the data may cause problems in estimating the completeness of death registration.

Techniques for Indirect Estimation of Mortality

As noted in earlier sections, registration of deaths is still incomplete in most developing countries. Even in countries where completeness is improving in adult ages, underregistration of infant and child deaths is still a serious problem. In cases where the data are too incomplete even to apply the evaluation and adjustment procedures presented in the last section, it is necessary to estimate mortality indirectly. This section presents some methods for indirect estimation of mortality.

Techniques for estimating mortality from age distributions

There are several techniques for estimating mortality indirectly from information on the age structure of the population. Five such techniques are presented here. The first two require that two censuses be available; these techniques involve the use of model life tables, a topic discussed in detail in a later section. The third may be used when only one census is available and it relies on stable population theory. The last two techniques use populations from two consecutive censuses by 5-year age groups, without using model life tables; they do not require that the censuses be a specific number of years apart.

When information on the population by age groups is available for two consecutive censuses, it is possible to calculate how many persons enumerated in the first census have survived up to the second one. This comparison is highly facilitated if the two census dates are 5 or 10 years apart. Then, it is easily understood that persons age 20 to 24 years enumerated in the first census would be 25 to 29 years old 5 years later or 30 to 34 years old 10 years later. If the censuses are not 5 or 10 years apart, the comparison is more cumbersome because the pertinent age groups would be in a nonconventional format. For example, if the censuses were 12 years apart, persons enumerated at ages 20 to 24 in the first census would be 32 to 36 years old in the second census. Instead of adjusting the population to such unusual age groups, it is recommended to interpolate (or extrapolate) the census information to make it refer to an interval of 5 or 10 years. The interpolation (or extrapolation) is done by choosing as pivotal points the same age groups in each census population.

Intercensal survival ratio technique. This is a straightforward procedure developed long ago to estimate the general level of mortality by calculating survival ratios for each cohort during an intercensal period from population distributions by age and sex from two censuses or surveys. The technique requires the following assumptions:
(1) That the completeness of the population information in all age groups is the same at both dates.

(2) That there was no migration during the intercensal period.

(3) That there was no age misreporting in either census.

(4) That the mortality pattern during the intercensal period is similar to a model pattern.

Intercensal survival ratios are calculated for all 5-year age groups and the results are compared with survival ratios obtained from model life tables to infer a general level of mortality. For example, if two censuses were taken 10 years apart, a survival ratio would be calculated as the ratio of the population at ages 30 to 34 years in the second census to the population ages 20 to 24 years in the first census. Both the actual intercensal survival ratios and the corresponding information from model life tables can be plotted on a graph to make a visual comparison of the mortality levels from the two sources.

Although intuitively "accurate," this technique has several problems resulting from the assumptions upon which it is based: many developing countries do experience international migration, age misreporting and incomplete census enumeration.

If there has been immigration, the technique would tend to underestimate mortality; in cases of high immigration in specific ages, the intercensal survival ratios could even reach values above one. If there has been emigration, on the other hand, the technique would overestimate mortality because persons leaving the country during the intercensal period would have the same effect on the calculations as persons who die.

Errors in age reporting can produce large variations in the calculated survival ratios, making it difficult to interpret the results. Finally, if the two censuses had different degrees of completeness of enumeration, the mortality estimate would be biased. In a specific example illustrated in figure III-7, survival ratios were taken from two consecutive censuses without smoothing and compared to survival ratios obtained from model life tables. It is difficult to determine the general level of mortality from the results. Furthermore, if the population of the more recent census were increased by 2 percent (a plausible estimate of underenumeration), revised survival ratios would indicate a much lower mortality level (that is, a higher life expectancy at birth). When the same calculations were made after smoothing the populations, the results were better, but it still remains difficult to estimate the level of mortality (see figure III-8). Because of such discrepancies between the assumptions and real situations, this technique is not frequently used, although it has been improved with the introduction of projection techniques and the use of computers, as explained below.
Figure III-7. Mexican and Model Survival Ratios Based on Reported Populations

Note: Model survival ratios pertain to the Coale-Demeny model life tables, Region West. Mexican ratios are from the 1940 and 1950 census populations.

Figure III-8. Mexican and Model Survival Ratios Based on Smoothed Populations

Note: Model survival ratios pertain to the Coale-Demeny model life tables, Region West. Mexican ratios are from the 1940 and 1950 smoothed census populations.
Projected survivors technique. This technique was developed as an improvement over the previous one to reduce the impact of age misreporting (United Nations, 1967). Although it has the same basic idea as the intercensal survival ratio technique, it compares a census population with a projected one, using most of the age groups. Like the previous technique, its objective is to estimate the general level of mortality during an intercensal period, and it requires an age distribution from two censuses or surveys.

The assumptions relating to this method are as follows:

1. Completeness of the population information in all age groups is the same at both dates.
2. There was no migration during the intercensal period.
3. There was no age misreporting (but the impact of a violation of this assumption is less severe than in the intercensal survival ratio technique).
4. The mortality pattern during the intercensal period is similar to a model pattern.

This procedure differs from the previous one in one main respect: instead of calculating survival ratios for the intercensal period, the earlier census is projected up to the date of the second census, and the projected population is compared with the enumerated population for the same date. Assuming the comparison involves two censuses 10 years apart, the following steps are carried out:

1. The population of the first census is projected by 5-year age groups to the date of the second one, using a set of survival ratios from model life tables pertaining to a "guessed" level of mortality. Since no fertility is used, the projected population refers only to ages 10 years and over.
2. The total projected population is compared with the total enumerated population age 10 years and over in the second census.
3. If the projected population differs from the enumerated, an iterative process begins. By generating model life tables, the mortality is increased (or decreased) if the projected population is greater (or smaller) than the enumerated population. The iterative process stops when both populations are the same.
4. The level of mortality that makes the projected population equal to the one enumerated in the second census is kept as one estimate of the mortality level.
5. The population of the first census is projected again by 5-year age groups, this time using only the population ages 5 years and above, thus disregarding the first age group, 0 to 4 years. The
projected population is compared with the total enumerated population ages 15 years and above.

(6) The procedures described in (3) and (4) are repeated and another level of mortality is obtained. This process is repeated for the population ages 10 and above, 15 and above, and so forth, in the first census. Usually, more than 10 projections are compared for 10 or more open-ended age groups, and hence more than 10 mortality levels are estimated. In each case, the result is influenced by age misreporting in only the lower bound of the open-ended age group (for example, the "15" in "age 15 and above"), and so the overall effect of age misreporting is much less than in the previous technique.

After all the calculations are complete, an analysis should be made of the mortality level estimated from each population projection to provide some guidance in selecting a final estimate. One suggestion is to take the median value of the life expectancies estimated in the first nine projections (referring to all ages, age 5 and over, up to age 40 and over) as representing the average mortality level during the intercensal period. Once such a life expectancy is accepted as representing the mortality level, a corresponding model life table can be generated with the United Nations program MATCH of the MORTPAK package.

Caution is advised in adopting the resulting mortality estimate, as a particular population's characteristics may not correspond to the assumptions implicit in the technique. In general, the greater the departure of the population's own characteristics from the technique's assumptions, the greater the bias in the final estimate.

This technique has the inconvenience of requiring the census populations for an exact number of years apart (5, 10 or 15 years). This implies that most census populations have to be shifted to specific dates. In addition, the iterative process of the methodology implies long calculations. Fortunately, other techniques have been developed that give similar results and are easier to use, such as the Preston-Bennett, Bennett-Horiuchi, and integrated techniques; these techniques are presented in this chapter. The Bennett-Horiuchi technique requires the intercensal deaths by age (or an estimate) and the integrated technique uses a pattern of mortality by age.

**Stable population estimates.** Even in countries with only one census, the age distribution may be used to estimate a general mortality level. In such a case, the actual age structure of the population is compared with the age structure of a stable population. As noted in an earlier section, a stable population is a theoretical population in which mortality and fertility have not changed for a long period of time, and in which there has been no migration. These conditions produce a population whose age distribution does not change either.

Age distributions of stable populations differ according to the mortality and fertility levels they have maintained. Thus, in order to match an actual population to a particular stable population, it is necessary to
have a notion of the level of mortality or fertility. But these levels are unknown, since the purpose of the technique is to estimate them. The growth rate of the population is taken as a possible substitute for the levels of mortality and fertility.

To estimate the general level of mortality from the age structure of a population at a single date, then, it is necessary to have an estimate of the rate of natural increase (which equals the population growth rate when there is no migration). Consistent with the requirements of a stable population, this technique assumes that the population's fertility has not changed during the past; that mortality has not changed during the past or has declined slowly; and that there has been no migration. It also assumes that there has been no differential completeness of enumeration by age groups and no age misreporting.

The technique consists of calculating the proportion of the population younger than a given age and matching it, by an iterative process, with a Coale-Demeny stable population that has the same proportion of persons under age x and the same growth rate as the real population. Once the stable population has been selected, the life expectancy at birth from the life table corresponding to the particular stable population is accepted as one estimated mortality level for the country. (A corresponding fertility level can also be determined in the process.) The procedure is repeated for several age groups of the population from birth to age x, and a final level of life expectancy is selected from the alternatives to serve as the mortality estimate for the country.

As in the case of the previous technique, it is advisable to select the median level of the nine life expectancies at birth pertaining to the youngest nine open-ended age groups as representing the mortality level of the population. A model life table can be generated with such a life expectancy at birth. The results of this technique are generally less robust than those derived from two censuses. The results are sensitive to the estimated rate of natural increase, and hence small errors in this rate can produce a significant impact on the estimated level of mortality. Other factors that may bias the estimates are age misreporting and international migration.

**Integrated technique.** This technique was developed for the purpose of estimating mortality levels from two consecutive population age structures (Preston, 1983). In addition to the level of mortality measured by life expectancies, the technique provides estimates of the crude birth rate. The age structure, level of mortality, and crude birth rates are "integrated" in the technique, and from this it derives its name. The development of the method uses logits and the "generalized" stable population equation (Preston and Coale, 1982). The technique also provides a "correct" intercensal population age structure.

---

1This "generalized" equation is similar to the "general" stable population equation, except that the intrinsic growth rate is not necessarily constant for all ages.
The integrated technique does not require any specific interval of years between the two age structures of the population. However, the assumptions are that the population has not had intercensal migration, that age misreporting is not significant, and that both censuses have the same coverage. In addition, it requires a model life table to be used as a standard in the logit calculations, and survivorship values from birth to age 1 and from birth to age 5. In populations where census errors and age misreporting are significant, the results may be biased (see appendix III-8). A computer program (PRESTO in the MORTPAK package) is available for the application of this technique.

Preston-Bennett technique. Another technique that estimates the general level of mortality based on two population age distributions was developed in the early 1980’s by Preston and Bennett (1983). The only data required for the application of this technique are the age structures from two censuses.

The objective is to estimate levels of life expectancy at ages 5 and above. The technique assumes that completeness of enumeration and possible age misreporting were the same in the two censuses, and that there was no international migration during the intercensal period.

Although the Preston-Bennett technique is based on relationships developed by stable population theory, it does not require that the population be stable. Based on the population enumerated in each age group and the average annual intercensal growth rate of the population in each age group, this technique estimates the cumulative number of years lived by the population and the number of persons at exact ages. From these estimates, life expectancies are calculated for each age 5 years and above.

This technique has several advantages. No information is required beyond the two censuses, and they do not have to be 5 or 10 years apart. In addition, no model life tables are used. Major disadvantages, however, are that the technique does not provide a life expectancy at birth, an estimate of infant mortality, nor a pattern of mortality by age. Furthermore, if the census data contain errors, or if there were differences in age misreporting or completeness of enumeration between the two censuses, or if the country experienced international migration during the intercensal period, then the set of life expectancies for the various ages may not follow an expected smooth and declining trend with age. Although the authors have suggested possible corrections of the results if the population has been exposed to migration or if the censuses have different degrees of completeness, it is difficult to carry out such adjustments in real cases (see appendix III-9 and software PREBEN in volume II).

A special caution

Although each of the mortality estimation techniques based on the population age structure may be appropriate under certain conditions for the purpose of estimating a general mortality level, none of them provide a reliable estimate of infant and child mortality. Although the microcomputer
programs associated with some of the techniques provide a life table with death rates for all ages, including infant mortality, users should realize that the life table is a model which does not necessarily reflect the mortality pattern and conditions of the actual population. Thus, special caution is required before accepting levels of infant and child mortality as represented by these programs. The next section provides some techniques for the indirect estimation of infant and child mortality.

**Techniques for indirect estimation of infant and child mortality**

The last section presented some useful techniques for the indirect estimation of the general level of mortality. However, these techniques do not provide estimates of one of the most useful mortality indices: the infant mortality rate. Various techniques have been developed to estimate mortality during the first years of life. These techniques use data on children ever born and children surviving by age of mother (or by duration of marriage/union): if it is known how many children a mother has had and how many of them are still alive, then the ratio of the children who have died to the total number born is a measure of mortality from birth up to a certain age. This concept is reflected in the Brass technique and four modifications of it, all of which produce estimates of the probability of dying from birth to early ages of childhood. Some general discussion of the methodology is appropriate before the specific techniques are presented.

Theoretically, estimates of child mortality could be transformed into probabilities of dying from birth to early ages of childhood, and these probabilities could be used to obtain the number of survivors from a given number of births (for example, 100,000) up to ages 1, 2, 3, 5, 10, 15, and 20. These survivors would be equivalent to the life table function $l_x$, representing the mortality conditions of the population during the recent past. However, in most actual cases, they will not appear in a consistent life table pattern because: (a) mortality and fertility may have changed during the past; (b) actual fertility and mortality patterns may be different from those used in developing the techniques; and (c) most importantly, there may be errors in the data. Moreover, the estimates will pertain to different periods of time in the past and hence will not directly represent the survivors at exact ages as they do in a life table.

The techniques can be used to estimate the levels of infant and child mortality. This is done using not necessarily the information for ages 15 to 19, as the techniques recommend, but for older age groups that give an estimate of the probability of dying from birth to specific ages. Since there is a rather strong relationship between mortality rates in infancy and those at subsequent ages, model life tables may be used to estimate infant mortality from mortality estimates pertaining to older ages. This is done by a matching process. For example, the Brass technique provides an estimate of the probability of surviving from birth to age 3 years based on information for the age group 25 to 29 years. A model life table may be identified that has the same mortality level as this probability of surviving to age 3, and the infant mortality rate from that model may be accepted as appropriate for the population. The corresponding life expectancy at birth may also be

101
accepted. The process of matching real data with the model is carried out automatically by the microcomputer programs presented in appendix III-10.

The above steps may be repeated using information pertaining to other age groups of the female population, and a different infant mortality rate estimated corresponding to each age group. All of these estimates may be valid, referring to mortality at different time periods. The older the mothers, the longer their children have been exposed to the risk of dying, and hence the more remote the year to which the estimates pertain. The reference year can be calculated by a procedure incorporated into the microcomputer program used for applying the methods (see appendix III-10).

As with other techniques, the quality of the data strongly affects the results. Unfortunately, it is difficult to evaluate information on the proportion of children who have died, by age of mother. However, data quality may be considered suspect if the estimates of child mortality fluctuate in an erratic manner. In most developing countries, mortality has been at higher or constant levels during the past, so any estimate suggesting that mortality has increased should be carefully evaluated.

Finally, these four techniques may be affected by significant changes in the pattern of fertility, especially if the new pattern differs considerably from patterns that were used for developing the techniques. In addition, a rapid and significant decline in mortality may go undetected by these techniques.

On the positive side, there is a recognized advantage in the use of these techniques in demographic analysis: they provide the possibility of determining mortality differentials by regions or special groups within the population. If the quality of information from censuses and/or surveys does not vary significantly by social groups or regions within the country, the application of these techniques can detect infant and child mortality differentials. Details of the five techniques are presented below.

**Brass technique.** This technique provides a method to estimate the level of infant and child mortality based on data on the average number of children ever born and the average number of children surviving per woman by age groups of the female population (Brass, et al., 1968). It assumes that mortality and fertility have not changed during a period of about 10 years prior to the reference date of the information and that there has been no age misreporting of women.

Since censuses or surveys generally do not ask the birthday or death day of the child, the technique estimates the average age of children who have died and transforms this information into probabilities of dying from birth to exact ages by using a simulation model. In the model, women are "simulated" to have children; these children are "survived," taking into account the ages of those who die. The simulation model uses: (a) a third degree polynomial to represent age-specific fertility rates, allowing the shape of the polynomial (fertility rates) to be modified by changing the mean age of mothers; and (b) mortality from the Brass (Brass, et al., 1968) general standard model.
Based on the simulation model, a set of adjustment factors has been developed to transform the proportion of children who have died into probabilities of dying from birth to exact ages, using age-specific fertility rates, infant and child mortality rates, and age of mother. These adjustment factors transform the information provided by mothers at age 15 to 19 years into an estimate of the probability of dying from birth to exact age 1. Similarly, the information provided by mothers age 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, and 45 to 49 is transformed into estimates of probabilities of dying from birth to ages 2, 3, 5, 10, 15 and 20, respectively.

As with other techniques, the results of the Brass technique may contain certain biases when the population characteristics depart from the assumptions made in developing the technique. Because the method assumes that mortality and fertility have not changed significantly during the past, an abrupt or large reduction in mortality or a significant change in the fertility level or pattern in the actual population may have a large effect on the estimates. Similarly, age misreporting of mothers may affect the results. But probably the most important factor affecting the estimates is the quality of the base information. It is not unusual for mothers to fail to report children who have died, especially those who died at a very early age. The underreporting of children who died produces an underestimation of infant and child mortality.

Generally, child mortality estimates are based on information on age groups of mothers 20 to 24 years and above. Theoretically, this technique estimates infant mortality using information from the age group 15 to 19 years, but it tends to overestimate infant mortality for the whole country. Women in this youngest age group may have socioeconomic characteristics significantly different from those of the total number of mothers whose infants die during a year. They also have a smaller proportion married and a smaller number of births than women in older age groups, which can result in more random error in the estimates. In addition, the model life tables often do not fit as well with the actual data at the youngest ages, causing possible further bias in the results (see appendix III-10). There is a computer program for calculating this technique developed by the United Nations, CEBCS in the MORTPAK package.

**Trussell and Sullivan techniques.** Two slight modifications have been made to the technique developed by Brass: one by Trussell (United Nations, 1983) and the other by Sullivan (1972). The modifications do not concern the assumptions nor the methodology; they relate only to the models of mortality and fertility used in the simulation model that calculates the factors transforming the proportion of children who died into probabilities of dying. The Trussell modification uses Coale-Trussell fertility models instead of the third degree polynomial used by Brass, and it uses mortality patterns from the Coale-Demeny model life tables instead of the "general standard model" used by Brass. The Sullivan modification uses patterns from actual data for fertility and Coale-Demeny models for mortality. These two modifications yield results similar to those of the original procedure presented by Brass. For practical purposes, the quality of data is more important than the modifications.

103
The objectives, assumptions, data requirements, and descriptions of the
techniques for these two modifications are the same as for the Brass technique
and hence are not repeated here (see appendix III-10).

**Feeney technique.** Feeney (1976, 1980) also modified the technique
developed by Brass. His modification pertains to the assumption concerning
mortality during the past. While Brass assumed that mortality has been
constant, Feeney assumes that it has been declining linearly. In addition,
Feeney developed a procedure that automatically provides estimates of the
dates to which the infant mortality estimates pertain. Otherwise, the Feeney
and Brass techniques are the same.

**Palloni-Heligman technique.** This technique uses the same basic concepts
as the technique originally developed by Brass (Palloni and Heligman, 1985).
Based on information on children ever born and children surviving, classified
by age of their mothers or by marriage duration, the authors developed a new
set of correlation coefficients. These coefficients transform the proportion
of dead children (in relation to those born) into probabilities of surviving
from birth to an exact age (1, 2, 3, 5, 10, 15, and 20 years). The
correlation coefficients were obtained from information derived from mortality
and fertility patterns. These patterns were:

2. Fertility patterns from model fertility schedules (Coale and
   Trussell, 1974).

The assumptions made in developing this new set of correlation
coefficients were:

1. Mortality and fertility were constant during the past.
2. There is no relationship between mortality of mothers and
   mortality of their children; between age of mother and parity; nor
   between age of mother and child mortality.
3. Age of mothers is reported correctly.
4. Completeness of reporting is the same for children ever born as
   for children surviving.
5. As the authors mention, the mortality model used has been well
   identified.

The advantage of this technique over the others is that the authors used
the United Nations model life tables that may better fit the mortality pattern
of some countries (see appendix III-10, and the United Nations computer
program CEBCS in the MORTPAK package).
Johnson technique. This technique is based on the relationship between two life table functions: the $l_0$ and the $L_0$. In a real population, the census population under age 1 year can be assumed to be similar to the $L_0$ function of the life table, while births during the year prior to the census can be taken as the $l_0$ function (U.S. Bureau of the Census, 1982). An index representing a significant part of the infant mortality rate can be obtained by taking the ratio of the population under age 1 year to births during the last year, and subtracting the ratio from one. In order to obtain the whole infant mortality rate, the author used separation factors for infant deaths derived from the relationships developed by Coale and Demeny.

The Johnson technique has the advantage that it does not require the use of model life tables to obtain the estimate. In addition, if the questions are asked, (a) "Did you have a birth during the last 12 months?" and, if yes, (b) "Is the baby still alive?" then the infant mortality estimates can be based on any group of women for whom the data are tabulated. However, the results may be affected by errors in reporting the number of births occurring during the year prior to the census or survey date and/or the number surviving. Also, the technique is affected if the reference period is not 1 year (see appendix III-11, and the spreadsheet BTHSRV in volume II).

Mortality under 5 years of age. The estimation of infant mortality based on information on the average number of children ever born and children surviving by age of women, has two aspects that need to be discussed. First, infant mortality theoretically can be accepted directly from information on women aged 15 to 19 years. However, such estimates are frequently biased because of the uniqueness of mothers at that particular age of the female population. Second, infant mortality rates estimated from other age groups are the result of a matching process between results of the technique and model life tables.

The first aspect may produce results that are not representative of the whole population. The second aspect produces estimates that depend on the model used. Both aspects have no theoretical solutions, mainly because the actual level of infant mortality is unknown, and model patterns of mortality have to be accepted as pertaining to the population.

Mortality rates for particular ages may fluctuate from country to country and from model to model at similar general levels of mortality (measured by the life expectancy at birth). For instance, one of the distinctive characteristics of the models is the different levels of infant mortality (and hence mortality from 1 to 4 years) for the same levels of life expectancies at birth. But, the mortality under age 5 years seems to fluctuate less. Based on mortality models (and hence on empirical life tables used for constructing them) for the same levels of total mortality, the probability of dying from birth to age 5 years fluctuates less than the infant mortality rate or the mortality from age 1 to 4 years.

Consequently, estimates of mortality from age 0 to age 5 years are more stable than infant mortality. Therefore, the mortality under age 5 years
should provide better trends of child mortality than those based on infant mortality.

Just as mortality estimates from birth to different ages are matched to produce estimates of infant mortality, the mortality under age 5 years also can be matched. In other words, the results obtained from the techniques presented previously for estimating infant mortality can be used for estimating mortality under age 5 years. The United Nations has developed software to obtain the estimates of mortality under age 5 years, by using the same information as for estimating infant mortality. This software is called QFIVE (United Nations, 1990).

Techniques for indirect estimation of adult mortality

Like mortality in infancy and childhood, mortality in adult ages can be estimated indirectly when reliable data are not available to measure it directly. Two principal techniques have been developed to estimate adult mortality based on information collected in censuses or surveys on the number of persons whose mother or father has died (orphanhood) and on the number of persons whose first spouse has died (widowhood). Both techniques provide an estimate of survivorship levels between two adult ages for a period of time prior to the year of data collection. The actual reference period can be estimated, but only under certain assumptions.

Orphanhood technique. The concept underlying the orphanhood technique for estimating adult mortality is that the proportion of persons whose mother is alive can be used to derive an estimate of the survivorship of women (or men) from a given age in adulthood to a subsequent older age (United Nations, 1983). If a child of age x is an orphan, it means that sometime during a period of x years, the child’s mother has died. To apply the technique, census or survey data must be available on the proportion of respondents whose mother is still alive for each 5-year age group of the population, as well as the proportion of respondents who do not know this information. Other versions of the technique may be used when such data are available for more than one date (see appendix III-12).

The orphanhood technique assumes that mortality and fertility have not changed during the past; that there are no mortality differentials between large and small families; that the data contain no age misreporting; and that the data refer to the survivorship status of the respondent’s natural mother (not a steppmother).

This technique follows procedures similar to those used to estimate infant and child mortality: mortality and fertility models were used to create simulated populations with various proportions of orphans in each age group. Using these simulated populations, a set of weighting factors was developed to relate a life table probability of surviving between two ages (from age 25 to ages 35, 40, 45, etc.) to the proportion of persons whose mother is still alive. When the technique is used, these weighting factors are applied to the proportions orphaned in the actual population to obtain estimates of the corresponding survival ratios.
This technique was originally developed by Brass, using the following models: for mortality, the Brass general standard model, using logits to generate a life table; and for fertility, a third degree polynomial. In later versions of the technique, other researchers used different models, such as Coale-Demeny life table models and fertility models developed by Trussell and Coale. As in the case of infant and child mortality, the modifications give similar results in any practical case, and errors in the data have a much greater impact on the results than the use of different mortality and fertility models.

The orphanhood technique can be applied in reference to survivorship of either the mother or the father of the respondent, but the weighting factors and the base age for estimating the survivorship values are different in the two cases. When available, actual data on mean age of mothers or fathers should be used. If the mean age of fathers is not available, the authors suggest a procedure for its estimation, adding to the mean age of mothers 9 months plus the difference between the two ages at marriage. Age at marriage can usually be obtained from census data on the proportion of the population ever married. If actual data are not available on the mean age of either parent, a standard age of 25 years for mothers and 32.5 years for fathers may be used.

A problem in using the orphanhood technique is to determine the particular time period to which the results refer, as the estimate from each age group pertains to a different time in the past. Especially in situations of declining mortality, the time reference is important. Although the microcomputer program pertaining to this technique includes an estimation of the reference period, it should be accepted only as a rough approximation (see appendix III-12).

In addition, various factors may bias the results. The base information on survivorship of parents is collected from respondents who themselves are alive at the time of the census. If mortality of parents is related to the mortality of their children, the estimate could be too low because some of the parents who died would have no surviving children to respond to the census. An opposite type of bias is produced if the question about orphanhood is asked of all census respondents. In this case, mortality of parents with many surviving children may be overrepresented. In other situations, responses could be erroneous if children had been adopted by other families and declared their stepmother or stepfather as a natural parent. Finally, like most other techniques for indirect estimation of mortality, the results may be affected by changing mortality and fertility during the past and by age misreporting.

The United Nations program ORPHAN in the MORTPAK package performs the calculations of this technique.

Widowhood technique. This technique for estimating adult mortality has the same theoretical basis as the orphanhood technique, but it is based on information on the proportion of men and women whose spouse from the first marriage or union is still alive, by 5-year age groups (Hill, 1977). While the orphanhood technique uses the mean age of mothers (or fathers), the widowhood technique uses the singulate mean age at marriage, which is obtained from the distribution of the ever-married population by sex and 5-year age
groups. If information on age at marriage is not available, it is possible to use data on duration of marriage, that is, the number of years that have elapsed since first marriage. The methodology assumes that mortality has been constant during the past; that there is no relationship between wife’s and husband’s mortality; and that there has been no age misreporting.

The widowhood technique is based on the results of simulation models which were created to generate data on the proportion widowed for various values of the mean age at marriage for men and women, at various mortality levels. From these data, regressions were run to relate survivorship probabilities between two specific ages to the mean age at marriage for men and women and the proportion of one sex remaining not widowed in each 5-year age group of the population. When the technique is applied, actual data from a census on the percent widowed in each age group and the mean age at marriage for each sex are compared to the regression coefficients to select an appropriate estimate of the male and female probabilities of surviving between two specific ages.

Results of this technique are affected not only by any violation of the assumptions used in developing the method, but also by erroneous responses to data queries concerning the first spouse. Incorrect responses are frequent, especially when the respondent is a family member other than the individual about whom spouse survivorship is asked. A child, for example, may not know how many times his or her parents have been married or the survivorship status of a parent’s first spouse. Some false responses may be deliberate, even when the respondent is the identified individual. Because of such difficulties, both the orphanhood and the widowhood techniques often do not yield acceptable results. Even when they do, there is still the problem of establishing the reference period to which the estimates pertain, although this can be approximated by a procedure incorporated into the microcomputer programs used for applying the methods. The technique is presented in appendix III-12, and there is a computer program, WIDOW, in the MORTPAK package of the United Nations.

Model Life Tables and Their Use in Mortality Analysis

Model life tables have been constructed principally during the past four decades based on historical experience indicating that the pattern of mortality by age is highly predictable. Thus, model life tables are constructed using mathematical procedures and mortality patterns observed in life tables pertaining to actual populations with reliable data.

Model life tables are frequently used in demographic analysis for two main purposes: to estimate a probable current mortality level and to project the mortality pattern to some future date. Use of models to estimate a current mortality level was introduced in an earlier section on estimating infant mortality and a general level of mortality (life expectancy at birth). Their use in projecting mortality patterns will be dealt with in chapter VIII.

Life table functions are strongly related by age. Thus, if values for some of the functions are available, the rest can be estimated with some
degree of certainty using model life tables. For example, if estimates of mortality are available for some ages, levels for other ages can be interpolated using model life tables. Similarly, if mortality rates for an actual population are severely distorted because of errors in the data, these rates can be smoothed using models (see the section on smoothing death rates).

Model life tables can even be used to estimate a possible level of life expectancy at birth based on an estimated infant mortality rate; or conversely, a level of infant mortality can be estimated from an estimate of the life expectancy at birth. Without the availability of model life tables, most of the techniques presented in this chapter could not have been developed nor could they be applied to most populations. Nevertheless, the possibility of applying several models (with each one yielding different results), as well as the misuse of models, could result in a proliferation of mortality estimates that reflect only the arbitrary decisions of a particular user. Although life table models are useful, if not essential, in demographic analysis, it is necessary to understand their techniques to appreciate their limitations.

Abundant use of model life tables has become possible as a consequence of the development of microcomputer programs which can reproduce, at any specified age, a desired level of central mortality rate, probability of dying, number of survivors from birth up to a specific age, or life expectancy (see appendix III-13 and the United Nations program MATCH in the MORTPAK package).

The first models were developed by the United Nations (1956) in the 1950's based on empirical life tables available for various developed countries at that time. These models were constructed by calculating correlations between probabilities of dying pertaining to two consecutive age groups. Starting with an arbitrary level of infant mortality, the rest of the probabilities of dying were obtained using the estimated regression equations. These early United Nations models have been superseded by other models (including a recent set by the United Nations), with improvements in two different areas: in the empirical data used to develop the models and in the methodology. These other models are discussed below.

**Coale-Demeny models**

A set of model life tables was developed by Coale and Demeny (1968) in the 1960's. These models use more information than the early United Nations models and classify the life tables into four different sets, labeled West, East, North, and South, according to the patterns of mortality in the predominant regions of Europe represented in the original data. (It should be noted that these regional designations do not refer to major world regions as is sometimes supposed.) In each of these sets, the life expectancy at age 10 was correlated with the probability of dying at different ages. Two functions were used, according to the level of life expectancy at birth. Once the probabilities of dying were obtained, all life table functions were calculated. In addition to the life tables, corresponding stable populations were derived, based on the model life tables and age-specific fertility rates.
for various levels of the total fertility rate (see appendix III-13 and the United Nations program MATCH in the MORTPAK package).

**General standard model**

At about the same time as the Coale-Demeny models appeared, Brass (Brass, et al., 1968) developed the General Standard Model. In this model, life tables are related by a two-parameter logit system, based on a general pattern of mortality and a linear relationship between two sets of logits. By changing the two parameters of the linear relationship, the general pattern is changed in relation to the level and pattern of mortality.

**United Nations models**

Most of the empirical life tables used in constructing the models described above pertained to developed countries. More recently, the United Nations published a new set of life table models, based on the mortality experience of developing countries with reliable information (Heligman, 1984 and United Nations, 1982). Empirical life tables for developing countries were grouped into four sets, plus a general set including all of them. These sets refer to the Latin American, Chilean, South Asian, Far East Asian and General models. The methodology, developed by Heligman, is based on the logit of the probabilities of dying at each age group when the level of mortality changes. The pattern of mortality in these models varies slightly (see figure III-9, appendix III-13 and the United Nations program MATCH in the MORTPAK package).

Other less frequently used model life tables are the Ledermann (1969) and the OECD (Organization for Economic Cooperation and Development) (1980) models.
The Use of Logits in Life Table Construction

Logits were developed for transforming the logistic function (usually known as the "S" curve) into a straight line. When a logit is applied to a function that varies between 0 and 1, the extreme values (near 0 and near 1) are "stretched out" while the middle values (near 0.5) are not changed very much. The logit function thus performs like a logarithmic function at the extremes and like a linear function in the middle of the range.

The logit of a value is calculated by subtracting the value from 1, dividing the difference by the value itself, and then taking the natural logarithm of the result. Sometimes, the result of this calculation is divided by 2 (Brass, 1975; United Nations, 1982 and 1983).

In symbols:

\[
Y = \text{logit } y = \frac{1}{2} \ln \frac{1 - y}{y}
\]

Where:

\(Y\) represents the logit of the function \(y\).
Several life table functions vary from 0 to 1, or they can be transformed so that they will vary from 0 to 1. The logits of life table functions such as $nQ_x$, $m_x$, and $l_x$ are still not linear functions of age, but the variations, especially at the extremes, are easier to observe as logits (see figures III-10-A and III-10-B). In addition, the relationship between the logits of these functions from two different life tables tends to be almost linear (see figures III-11-A and III-11-B). The assumption of linearity permits the use of the logit in mortality analysis.

In symbols:

$$Y_1 = a + b Y_2$$

Where:

$Y_1$ and $Y_2$ represent the logits of a particular function of life tables 1 and 2, respectively; and

$a$ and $b$ are constants of the linear relationship.

In practice, "life table 2" in the above calculations would be referred to as the "standard life table," which in most cases would be a model, an earlier life table from the same country, or a life table from a neighboring or otherwise similar country. In demographic analysis, logits are used for two main purposes: (a) for smoothing life table functions; and (b) for constructing a life table from fragmentary information on mortality.

**Smoothing with logits**

Logits are used to smooth life table functions when information is so distorted that other smoothing techniques do not give acceptable results, for example when information on deaths collected in a survey reflects severe age misreporting and/or underreporting of deaths in certain ages. In such cases, the distorted life table function derived from the reported data can be smoothed and even adjusted by using logits, with another existing life table taken as the standard. The steps are as follows:

1. Take logits of the selected function for both life tables.
2. Calculate constants of the linear relationship ($a$ and $b$ in the above formula) on the two sets of logits using one of various methods (for example, least squares, grouped means, or trimmed means).
Figure III-10. Logits of Selected Life Table Functions

**A**  
Logits of $I_x$ Function

**B**  
Logits of $q_{n.x}$ Function
Figure III-11. Relationship Between the Logits of Two Selected Life Table Functions

Note: The standard and actual logits pertain to Coale-Demeny and United Nations model life tables, respectively, for females with $E(o) = 55$. 

Note: The standard and actual logits pertain to Coale-Demeny and United Nations model life tables, respectively, for females with $E(o) = 55$. 

114
(3) Use constants a and b and the logits of the standard life table to obtain new smoothed logit values for the population under study.

(4) Take anti-logits of the smoothed logits to obtain the smoothed life table function.

This procedure provides nicely smoothed values of the function. However, these values should be used with caution. The pattern of mortality in the smoothed values will resemble the pattern of the standard life table and not necessarily the pattern of the population being studied. Furthermore, the selection of an alternative life table function for smoothing might yield quite different results. For example, if the \( nq_x \) function is smoothed and a corresponding life table calculated, the result may vary considerably from another life table calculated based on the same procedure but smoothing the \( l_x \) values. There are no guidelines available to determine which function to select for smoothing. Also, the smoothed values of any of the functions could imply levels of mortality for certain ages (mainly among infants and children) so different from those expected or observed that the smoothed values cannot be accepted. See figure III-12 for an example of using logits for smoothing. Bureau of the Census spreadsheets LOGITLX and LOGITQX performing these calculations are presented in appendix III-14, and their documentation in volume II.

**Figure III-12. Comparison of Probabilities of Dying: Empirical, Model, and Smoothed**

- **Rates**
- **Age**

Note: The model pertains to the Coale-Demeny model life tables, Region West, females with \( E(o) \geq 50 \). Empirical rates have \( E(o) \geq 50.3 \) and logits were used in smoothing the rates.
In some cases, it is known before smoothing that mortality information for certain ages is useless, and such information should not be used in the smoothing process. In this case, it is advisable to use the fragmentary information of mortality with the United Nations Program, BESTFT, of the MORTPAK package (see next section and appendix III-14). The user should determine whether the results are reasonable before accepting them as estimates.

**Constructing a life table from fragmentary data**

The logit procedure for constructing a life table does not differ greatly from the smoothing procedure described above, although it is more straightforward. An appropriate application is a simple case in which an infant mortality rate is available and a corresponding life table is needed. To apply the technique, a standard life table is required, but instead of calculating the values of a and b as above, b is assumed to equal 1. Using the logit corresponding to the observed infant mortality rate and the logit of the standard life table, the constant "a" is calculated (see above box), and the logits for other ages are based on the standard life table. Next the anti-logits are taken and then the life table is constructed (see appendix III-14).

In another case, if more than one mortality rate is available (for example, for several age groups), a similar procedure can be applied to obtain a life table.

As for other techniques, use of computers facilitates the calculations. The latest computer programs developed by Heligman at the United Nations not only simplify the procedures, but also improve on the technique. The logit procedure explained above is similar to the "one component" logit procedure used in the United Nations program BESTFT. The programs provide two additional procedures: the "two component" and the "three component" logit systems. If fragmentary data are available for more than one age group, results from the two or three components procedures are theoretically better than the one component procedure. The two and three components procedures take into account more information, but they also require better quality in the information used (see appendix III-14 and program BESTFT in the MORTPAK package (Heligman, 1984; and United Nations, 1982)).

A computer program developed by the United Nations (MATCH from MORTPAK) may be used for estimating life tables in areas where the general level of mortality or a mortality rate for any specific age group is known (such as a life expectancy at birth, an infant mortality rate, or a mortality rate for ages x to x+5) and an age pattern of mortality is available that can be accepted as pertaining to the same area. The program provides a life table with the desired level of mortality (or specified mortality rate) and an age pattern of mortality derived from the accepted pattern. For example, based on a life table available for the whole country, urban and rural life tables could be estimated if, for each area: (a) infant mortality rates are available; and (b) the age pattern of mortality for the whole country is assumed to be appropriate. The MATCH program in the MORTPAK package adjusts the country's age pattern of mortality (which corresponds to the national
level of infant mortality) to the infant mortality level of each area and calculates a life table for each of the areas (see appendix III-14).
Appendix III-1

Standardization and Difference Between Two Rates

There are two procedures for standardizing crude death rates: direct and indirect (Spiegelman, 1968). Direct standardization not only provides a better comparison of rates than the indirect procedure, but it also provides for an analysis of the difference between two crude death rates.

Direct Standardization

The procedure and tables presented in the text for six populations can be illustrated based on just two populations, for example, populations A and F of table A-III-1.1, for which information on deaths is also available. The crude death rates of both populations are calculated by dividing the total deaths by the corresponding total populations, times 1,000. The rates are 10.40 and 12.70, for populations A and F, respectively. From these data, population F appears to have higher mortality than population A. However, if the age-specific mortality rates are calculated, the opposite is shown to be true (table A-III-1.2). All age-specific mortality rates of population F are lower than those of population A. The higher crude death rate of population F is the result of the effect of the age structure. Population F has an older age structure than population A, and hence a higher proportion of people in ages of high-risk mortality.

The direct standardization process eliminates the effect of the age structure among a group of populations by using a single age structure as a standard for all of them. In the example, the standard age structure can be taken from either A or F, or from any other population. Taking as a standard the age structure of population F, the standardized crude death rate of A can be obtained as follows (using data from table A-III-1.1):

1. To calculate the number of deaths, each age-specific mortality rate of population A (the $M_x$ values in table A-III-1.2, column A) is multiplied by the population of F in each corresponding age group (the $P_{x,A}$ values in table A-III-1.1, column F) and they are divided by 1,000.

\[ nD(F,A)_x = \frac{nP_{x,F} \times nM_{x,A}}{1,000} \]

For instance, for the age group 40 to 44 years old:

\[ 3,528 = \frac{488,000 \times 7.23}{1,000} \]

2. The deaths resulting from step (1), shown in table A-III-1.3, panel A, column A, are totaled to 136,441, symbolized by $D(F,A)$, and divided by the total of the standard population, in this case population F, 5,544,000, symbolized by $P_F$. The result is the crude death rate $d(f,a)$ with the age structure of population F and the mortality of population A.
\[ d(F, A) = 1,000 \times D(F, A) / P_f \]

\[ 24.61 = 1,000 \times 136,441 / 5,544,000 \]

(3) This result, 24.61, represents the standardized crude death rate of population A, using the population and age structure of F as a standard.

A comparison of the standardized crude death rate (24.61) and the actual crude death rate of F (12.71), both with the same population, indicates that mortality is higher in A than in F. However, it cannot be concluded that mortality is almost twice as high in A as in F. Standardized crude death rates permit only the ranking, not the measurement, of levels of mortality in different areas.

Table A-III-1.1. Selected Populations and Deaths, by 5-Year Age Groups

<table>
<thead>
<tr>
<th>Age</th>
<th>A) Populations in thousands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Total</td>
<td>16,492</td>
</tr>
<tr>
<td>Under 1</td>
<td>708</td>
</tr>
<tr>
<td>1-4</td>
<td>2,480</td>
</tr>
<tr>
<td>5-9</td>
<td>2,594</td>
</tr>
<tr>
<td>10-14</td>
<td>2,166</td>
</tr>
<tr>
<td>15-19</td>
<td>1,807</td>
</tr>
<tr>
<td>20-24</td>
<td>1,464</td>
</tr>
<tr>
<td>25-29</td>
<td>1,154</td>
</tr>
<tr>
<td>30-34</td>
<td>919</td>
</tr>
<tr>
<td>35-39</td>
<td>749</td>
</tr>
<tr>
<td>40-44</td>
<td>615</td>
</tr>
<tr>
<td>45-49</td>
<td>504</td>
</tr>
<tr>
<td>50-54</td>
<td>405</td>
</tr>
<tr>
<td>55-59</td>
<td>316</td>
</tr>
<tr>
<td>60-64</td>
<td>244</td>
</tr>
<tr>
<td>65-69</td>
<td>177</td>
</tr>
<tr>
<td>70-74</td>
<td>116</td>
</tr>
<tr>
<td>75-79</td>
<td>49</td>
</tr>
<tr>
<td>80+</td>
<td>25</td>
</tr>
</tbody>
</table>

120
### Table A-III-1.1. Selected Populations and Deaths, by 5-Year Age Groups (Continued)

<table>
<thead>
<tr>
<th>Age</th>
<th>A) Deaths in thousands</th>
<th>B) Deaths in thousands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>171,517</td>
<td>81,563</td>
</tr>
<tr>
<td>Under 1</td>
<td>61,938</td>
<td>16,409</td>
</tr>
<tr>
<td>1-4</td>
<td>22,552</td>
<td>4,489</td>
</tr>
<tr>
<td>5-9</td>
<td>8,239</td>
<td>1,983</td>
</tr>
<tr>
<td>10-14</td>
<td>3,958</td>
<td>1,418</td>
</tr>
<tr>
<td>15-19</td>
<td>5,028</td>
<td>1,982</td>
</tr>
<tr>
<td>20-24</td>
<td>5,605</td>
<td>1,865</td>
</tr>
<tr>
<td>25-29</td>
<td>4,901</td>
<td>1,520</td>
</tr>
<tr>
<td>30-34</td>
<td>4,467</td>
<td>1,516</td>
</tr>
<tr>
<td>35-39</td>
<td>4,340</td>
<td>1,585</td>
</tr>
<tr>
<td>40-44</td>
<td>4,449</td>
<td>1,876</td>
</tr>
<tr>
<td>45-49</td>
<td>4,751</td>
<td>2,341</td>
</tr>
<tr>
<td>50-54</td>
<td>5,355</td>
<td>3,279</td>
</tr>
<tr>
<td>55-59</td>
<td>5,906</td>
<td>4,629</td>
</tr>
<tr>
<td>60-64</td>
<td>6,840</td>
<td>6,672</td>
</tr>
<tr>
<td>65-69</td>
<td>7,437</td>
<td>8,294</td>
</tr>
<tr>
<td>70-74</td>
<td>7,575</td>
<td>9,978</td>
</tr>
<tr>
<td>75-79</td>
<td>4,984</td>
<td>5,508</td>
</tr>
<tr>
<td>80+</td>
<td>5,199</td>
<td>6,309</td>
</tr>
</tbody>
</table>

Crude death rates: 10.40, 7.64, 4.92, 7.37, 12.70

### Table A-III-1.2. Age-specific Mortality Rates for Selected Populations (per thousand)

<table>
<thead>
<tr>
<th>Age</th>
<th>Mortality rates for populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>87.48</td>
</tr>
<tr>
<td>1-4</td>
<td>9.09</td>
</tr>
<tr>
<td>5-9</td>
<td>2.40</td>
</tr>
<tr>
<td>10-14</td>
<td>1.83</td>
</tr>
<tr>
<td>15-19</td>
<td>2.70</td>
</tr>
<tr>
<td>20-24</td>
<td>3.83</td>
</tr>
<tr>
<td>25-29</td>
<td>4.25</td>
</tr>
<tr>
<td>30-34</td>
<td>4.86</td>
</tr>
<tr>
<td>35-39</td>
<td>5.79</td>
</tr>
<tr>
<td>40-44</td>
<td>7.23</td>
</tr>
<tr>
<td>45-49</td>
<td>9.43</td>
</tr>
<tr>
<td>50-54</td>
<td>13.22</td>
</tr>
<tr>
<td>55-59</td>
<td>18.69</td>
</tr>
<tr>
<td>60-64</td>
<td>28.03</td>
</tr>
<tr>
<td>65-69</td>
<td>42.02</td>
</tr>
<tr>
<td>70-74</td>
<td>65.30</td>
</tr>
<tr>
<td>75-79</td>
<td>101.71</td>
</tr>
<tr>
<td>80+</td>
<td>207.98</td>
</tr>
</tbody>
</table>

Source: Calculated from table A-III-1.1.
Table A-III-1.3. Standardized Deaths and Crude Death Rates: Selected Populations

### A) Population age structure of F as standard and mortality from each population

<table>
<thead>
<tr>
<th>Age</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>136,441</td>
<td>109,668</td>
<td>89,489</td>
<td>88,316</td>
<td>75,383</td>
<td>70,409</td>
</tr>
<tr>
<td>Under</td>
<td>3,849</td>
<td>2,117</td>
<td>1,463</td>
<td>3,097</td>
<td>959</td>
<td>822</td>
</tr>
<tr>
<td>1-4</td>
<td>1,528</td>
<td>632</td>
<td>454</td>
<td>830</td>
<td>208</td>
<td>160</td>
</tr>
<tr>
<td>5-9</td>
<td>488</td>
<td>244</td>
<td>24</td>
<td>160</td>
<td>116</td>
<td>92</td>
</tr>
<tr>
<td>10-14</td>
<td>459</td>
<td>235</td>
<td>203</td>
<td>154</td>
<td>117</td>
<td>96</td>
</tr>
<tr>
<td>15-19</td>
<td>682</td>
<td>380</td>
<td>341</td>
<td>216</td>
<td>226</td>
<td>195</td>
</tr>
<tr>
<td>20-24</td>
<td>1,003</td>
<td>568</td>
<td>519</td>
<td>327</td>
<td>351</td>
<td>304</td>
</tr>
<tr>
<td>25-29</td>
<td>1,291</td>
<td>723</td>
<td>651</td>
<td>427</td>
<td>438</td>
<td>379</td>
</tr>
<tr>
<td>30-34</td>
<td>1,444</td>
<td>1,095</td>
<td>944</td>
<td>699</td>
<td>641</td>
<td>556</td>
</tr>
<tr>
<td>35-39</td>
<td>2,178</td>
<td>1,616</td>
<td>1,295</td>
<td>1,003</td>
<td>887</td>
<td>772</td>
</tr>
<tr>
<td>40-44</td>
<td>3,530</td>
<td>2,211</td>
<td>1,747</td>
<td>1,458</td>
<td>1,252</td>
<td>1,105</td>
</tr>
<tr>
<td>45-49</td>
<td>4,101</td>
<td>2,813</td>
<td>2,040</td>
<td>1,851</td>
<td>1,528</td>
<td>1,368</td>
</tr>
<tr>
<td>50-54</td>
<td>5,567</td>
<td>4,025</td>
<td>3,048</td>
<td>2,775</td>
<td>2,413</td>
<td>2,198</td>
</tr>
<tr>
<td>55-59</td>
<td>7,457</td>
<td>5,718</td>
<td>3,926</td>
<td>3,942</td>
<td>3,142</td>
<td>2,873</td>
</tr>
<tr>
<td>60-64</td>
<td>10,849</td>
<td>7,234</td>
<td>5,944</td>
<td>6,043</td>
<td>5,027</td>
<td>4,639</td>
</tr>
<tr>
<td>65-69</td>
<td>15,083</td>
<td>11,407</td>
<td>9,618</td>
<td>9,158</td>
<td>7,797</td>
<td>7,247</td>
</tr>
<tr>
<td>70-74</td>
<td>26,039</td>
<td>22,745</td>
<td>18,440</td>
<td>18,184</td>
<td>15,960</td>
<td>15,032</td>
</tr>
<tr>
<td>75-79</td>
<td>30,157</td>
<td>26,905</td>
<td>25,574</td>
<td>25,197</td>
<td>23,364</td>
<td>22,331</td>
</tr>
<tr>
<td>80+</td>
<td>24.61</td>
<td>19.78</td>
<td>16.14</td>
<td>15.93</td>
<td>13.60</td>
<td>12.70</td>
</tr>
</tbody>
</table>

Standardized crude death rates:

<table>
<thead>
<tr>
<th>Age</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>171,517</td>
<td>103,705</td>
<td>76,581</td>
<td>104,537</td>
<td>55,622</td>
<td>49,188</td>
</tr>
<tr>
<td>Under</td>
<td>61,938</td>
<td>34,069</td>
<td>23,548</td>
<td>49,830</td>
<td>15,424</td>
<td>13,229</td>
</tr>
<tr>
<td>1-4</td>
<td>22,552</td>
<td>9,332</td>
<td>6,696</td>
<td>12,250</td>
<td>3,075</td>
<td>2,355</td>
</tr>
<tr>
<td>5-9</td>
<td>6,230</td>
<td>3,121</td>
<td>311</td>
<td>2,042</td>
<td>1,479</td>
<td>1,171</td>
</tr>
<tr>
<td>10-14</td>
<td>3,958</td>
<td>2,025</td>
<td>1,754</td>
<td>1,330</td>
<td>1,007</td>
<td>827</td>
</tr>
<tr>
<td>15-19</td>
<td>5,028</td>
<td>2,805</td>
<td>2,512</td>
<td>1,596</td>
<td>1,668</td>
<td>1,436</td>
</tr>
<tr>
<td>20-24</td>
<td>5,605</td>
<td>3,176</td>
<td>2,899</td>
<td>1,827</td>
<td>1,959</td>
<td>1,697</td>
</tr>
<tr>
<td>25-29</td>
<td>4,901</td>
<td>2,744</td>
<td>2,470</td>
<td>1,623</td>
<td>1,664</td>
<td>1,441</td>
</tr>
<tr>
<td>30-34</td>
<td>4,467</td>
<td>2,515</td>
<td>2,169</td>
<td>1,607</td>
<td>1,673</td>
<td>1,278</td>
</tr>
<tr>
<td>35-39</td>
<td>4,340</td>
<td>2,542</td>
<td>2,037</td>
<td>1,579</td>
<td>1,395</td>
<td>1,214</td>
</tr>
<tr>
<td>40-44</td>
<td>4,449</td>
<td>2,787</td>
<td>2,202</td>
<td>1,837</td>
<td>1,578</td>
<td>1,393</td>
</tr>
<tr>
<td>45-49</td>
<td>4,751</td>
<td>3,259</td>
<td>2,364</td>
<td>2,144</td>
<td>1,770</td>
<td>1,585</td>
</tr>
<tr>
<td>50-54</td>
<td>5,355</td>
<td>3,872</td>
<td>2,932</td>
<td>2,669</td>
<td>2,321</td>
<td>2,115</td>
</tr>
<tr>
<td>55-59</td>
<td>5,906</td>
<td>4,528</td>
<td>3,109</td>
<td>3,122</td>
<td>2,489</td>
<td>2,276</td>
</tr>
<tr>
<td>60-64</td>
<td>6,840</td>
<td>5,391</td>
<td>3,850</td>
<td>3,810</td>
<td>3,170</td>
<td>2,925</td>
</tr>
<tr>
<td>65-69</td>
<td>7,437</td>
<td>6,117</td>
<td>4,742</td>
<td>4,515</td>
<td>3,844</td>
<td>3,573</td>
</tr>
<tr>
<td>70-74</td>
<td>7,575</td>
<td>6,430</td>
<td>5,047</td>
<td>4,931</td>
<td>4,222</td>
<td>3,946</td>
</tr>
<tr>
<td>75-79</td>
<td>4,984</td>
<td>4,353</td>
<td>3,529</td>
<td>3,480</td>
<td>3,055</td>
<td>2,877</td>
</tr>
<tr>
<td>80+</td>
<td>5,199</td>
<td>4,639</td>
<td>4,609</td>
<td>4,344</td>
<td>4,028</td>
<td>3,850</td>
</tr>
</tbody>
</table>

Standardized crude death rates:

<table>
<thead>
<tr>
<th>Age</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>24.61</td>
<td>19.78</td>
<td>16.14</td>
<td>15.93</td>
<td>13.60</td>
<td>12.70</td>
</tr>
</tbody>
</table>

Source: Calculated from tables A-III-1.1 and A-III-1.2.
Differences Between Two Rates

The process of standardizing crude death rates facilitates the analysis of the difference between two crude death rates (Kitagawa, 1955). Specifically, the process permits the decomposition of the difference between two rates into the parts pertaining to: (a) age structure; (b) mortality; and (c) interaction (the combined effect of age structure and mortality).

Using the information presented for populations A and F in tables A-III-1.1, A-III-1.2 and A-III-1.3, the difference between the two crude death rates can be explained by the contribution of the different age structures, the different mortality levels, and the interaction of the two. The question to be answered is why population F has a higher crude death rate than population A. The following steps give the answer.

(1) Total difference. The total difference between the two crude death rates (table A-III-1.1) is:

\[ 12.70 - 10.40 = 2.30 \]

(2) Age structure contribution. The contribution of the age structure can be estimated by comparing the crude death rate of population A, represented by \( d(A) \), with a crude death rate calculated based on the age structure of population F and the mortality of population A, represented by \( D(F,A) \). Since the mortality of both populations is the same, any difference between them is due to the different age structures of the populations. Multiplying the mortality rates of population A (table A-III-1.2) by the age structure of population F (table A-III-1.1), deaths by age groups are obtained (table A-III-1.3, panel A, column A). The total of these deaths (136,441) times 1,000 divided by the total population of F (5,544,000, from table A-III-1.1) gives a crude death rate, represented by \( d(F,A) \), with the age structure of population F and the mortality of population A.

\[ d(F,A) = \frac{1,000 \times 136,441}{5,544,000} = 24.61 \]

Consequently, the difference between the crude death rate \( d(F,A) \) and the actual rate of population A--\( d(A) \)--represents the contribution of the different age structures of the populations to the difference between the crude death rates of populations A and F.

\[ d(F,A) - d(A) = 24.61 - 10.40 = 14.21 \]

In other words, if population A had the age structure of population F, the crude death rate would be 14.21 points higher.

(3) Mortality contribution. To estimate the contribution of the difference in mortality between populations A and F to the difference between their crude death rates, a procedure similar to that of step (1) is followed. The crude death rate of population
Let $d(A) = 10.40$ be compared with a crude death rate based on the age structure of $A$ and the mortality of $F$, $d(A,F)$, as obtained from tables A-III-1.1 and A-III-1.2, columns $A$ and $F$, respectively. Their product, the number of deaths, is presented in table A-III-1.3, panel B, column $F$. The total of column $F$ (49,188) represents the total number of deaths that population $A$ would have if it had the same mortality as population $F$. Finally, the crude death rate is calculated as the total number of deaths divided by the total population of $A$ (16,492,000, table A-III-1.1):

\[
d(A,F) = 1,000 \times 49,188 / 16,492,000 = 2.98
\]

The difference between this crude death rate and the crude death rate of population $A$ represents the contribution made by the difference between the mortality of populations $A$ and $F$.

\[
d(A,F) - d(A) = 2.98 - 10.40 = -7.42
\]

Thus, if population $A$ had the mortality of population $F$, the crude death rate would be 7.42 points lower.

(4) Interaction. Interaction is the contribution to the difference between the crude death rates of populations $A$ and $F$ that is not explained independently by the age structure and mortality. The difference between the crude death rates of $F$ and $A$ is

\[
d(F) - d(A) = 2.30
\]

The difference due to age structure is

\[
d(F,A) - d(A) = 14.21
\]

The difference due to mortality is

\[
d(A,F) - d(A) = -7.42
\]

The difference between the two rates is

\[
2.30 = 14.21 - 7.42 + I
\]

Hence, the interaction $I$ is:

\[
I = 2.30 - 14.21 + 7.42 = -4.49
\]

Thus, the combined effect of the different mortality and age structure between the two populations reduces the crude death rate by 4.49 points.

124
**Indirect Standardization**

Crude death rates for populations whose age composition is known but whose number of deaths is not available by age (and hence age-specific mortality rates cannot be calculated) can also be standardized to determine the relative differences between the mortality of such populations and that of a standard. In cases where no age-specific mortality rates exist, but crude death rates are available, the crude death rates can be indirectly standardized to indicate how the areas rank in relation to the level of mortality of the standard population. However, the indirectly standardized rates do not represent the actual level of mortality of each area.

Indirect standardization consists, first, of estimating a ratio between two crude death rates, representing the relationship of the mortality of a certain population to the mortality of a standard population. Second, the product of this ratio times the crude death rate of the standard population gives the standardized crude death rate for the specific population. The mentioned ratio can be calculated if age-specific mortality rates are available for the standard population. As an example, population F will be taken as the standard (from tables A-III-1.1 and A-III-1.2) and the crude death rate of population B will be indirectly standardized (assuming that population B has no information on age-specific mortality rates). The steps in this process are:

1. The standard age-specific mortality rates (MF<sub>x</sub> from table A-III-1.2, column F) are applied to the populations whose crude death rate is to be standardized, in this case population B, table A-III-1.1, column B. The products of the standard age-specific mortality rates (table A-III-1.2, column F) times the age structure of population B (table A-III-1.1, column B) gives the expected number of deaths in each age group under the assumption that the population has the same mortality as the standard (table A-III-1.4, column B).

\[ D(B,F)_x = MF_x \cdot PB_x \]

Where:

- \( MF_x \) represents the standard age-specific mortality rate of age \( x \) (or age group \( x \)); and
- \( PB_x \) represents the persons of age \( x \) (or age group \( x \)) of population B.

For example, for the age group 25 to 29, the expected number of deaths (798, table A-III-1.4, column B) is the product of the population of B in that age group (639,000, table A-III-1.1, column B) times the standard mortality for the same age group (1.25 per 1,000, table A-III-1.2, population F).

\[ 799 = 639,000 \times 1.25 / 1000 \]
(2) The sum of the deaths for all ages calculated in step (1)--42,789, table A-III-1.4, column B--are divided by the total population of B (10,681,000, table A-III-1.1). This ratio times 1,000 is the crude death rate of population B, under the mortality conditions of the standard population.

\[ d(A,F) = 1,000 \times \frac{D(A,F)}{PF} \]
\[ = 1,000 \times \frac{42,789}{10,681,000} = 4.01 \]

(3) A ratio (K) is calculated by dividing the actual crude death rate of population B--7.64, table A-III-1.1, panel B--by the crude death rate calculated in step (2), 4.01. If this ratio is over one (or lower), it indicates that mortality in the population is higher (or lower) than the mortality of the standard population (table A-III-1.4).

\[ K = \frac{d(B)}{d(B,F)} \]
\[ = \frac{7.64}{4.01} = 1.905 \]

(4) The indirectly standardized crude death rate for population B (table A-III-1.4) is obtained by multiplying the ratio K by the standard crude death rate d(F).

\[ sd(B) = K \times d(F) \]
\[ = 1.905 \times 12.70 = 24.19 \]

The result indicates that mortality in B is higher than in F.
### Table A-III-1.4. Indirect Standardization: Selected Populations

<table>
<thead>
<tr>
<th>Age</th>
<th>Number of deaths from mortality of population F as standard and age structure of each population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Total</td>
<td>49,188</td>
</tr>
<tr>
<td>Under 1</td>
<td>13,229</td>
</tr>
<tr>
<td>1-4</td>
<td>2,355</td>
</tr>
<tr>
<td>5-9</td>
<td>1,171</td>
</tr>
<tr>
<td>10-14</td>
<td>827</td>
</tr>
<tr>
<td>15-19</td>
<td>1,436</td>
</tr>
<tr>
<td>20-24</td>
<td>1,697</td>
</tr>
<tr>
<td>25-29</td>
<td>1,441</td>
</tr>
<tr>
<td>30-34</td>
<td>1,278</td>
</tr>
<tr>
<td>35-39</td>
<td>1,214</td>
</tr>
<tr>
<td>40-44</td>
<td>1,393</td>
</tr>
<tr>
<td>45-49</td>
<td>1,585</td>
</tr>
<tr>
<td>50-54</td>
<td>2,115</td>
</tr>
<tr>
<td>55-59</td>
<td>2,276</td>
</tr>
<tr>
<td>60-64</td>
<td>2,925</td>
</tr>
<tr>
<td>65-69</td>
<td>3,573</td>
</tr>
<tr>
<td>70-74</td>
<td>3,946</td>
</tr>
<tr>
<td>75-79</td>
<td>2,877</td>
</tr>
<tr>
<td>80+</td>
<td>3,850</td>
</tr>
</tbody>
</table>

(1) Crude death rates using standard mortality and corresponding population
2.98  4.01  3.35  3.85  8.27  12.70

(2) Ratios of actual crude death rates to rates with standard mortality
3.49  1.91  1.47  1.91  1.08

(3) Crude death rates indirectly standardized
44.28 24.21 18.65 24.31 13.67 12.70

Notes:

(1) Mortality rates from table A-III-1.2 column F, multiplied by the population age structures from table A-III-1.1, columns A to E. The total of each column of this table is divided by the total population from table A-III-1.1.

(2) Ratios of actual crude death rates from table A-III-1.1 (panel B) and those rates calculated in (1).

(3) Rates calculated by multiplying the crude death rate of the standard population F (12.70) by the ratios calculated in (2).
Appendix III-2

Construction of a Life Table

There are several procedures for calculating a life table. This appendix concentrates on the procedure explained in the text. A general formula relating the mortality information to the probabilities of dying was presented on page 77. In order to use this generalized formula (as well as the one for calculating the \( L_x \) function), separation factors are required.

This appendix first explains how to estimate separation factors. Then it considers the calculation of infant and child mortality rates and the open-ended age functions of the life table.

Separation Factors

A separation factor is defined as the average time lived during an age interval by persons who die between the beginning and end of the interval. For instance, those dying between ages 57 and 58 years live on the average \( \frac{1}{2} \) year between the celebration of their 57th birthday and their time of death before celebrating their 58th birthday. However, separation factors are usually calculated for 5-year age groups. In this case, persons dying between ages 35 and 40 years live on the average about 2.5 years: some of them die immediately after celebrating their 35th birthday, while others die just before reaching their 40th birthday.

A frequently accepted approximation of the values of the separation factors is half of the age interval for which the separation factors are being calculated. For instance, for constructing abridged life tables, separation factors for the 5-year age groups over age 5 years are accepted to be 2.5, except for theoretical refinements based on excellent death statistics by year of birth or in the case of life tables constructed using other procedures relating (through correlations) the values of central death rates and probabilities of dying.

Although it is accepted that separation factors for ages 5 years and over are half of the age interval (when the age interval is not greater than 5 years), this assumption is not accepted for ages under 5 years. Separation factors for ages under 1 year and from 1 to 4 years are smaller than half the age interval. This is due to the fact that mortality is very high during the first day of life and declines rapidly during the first year and up to the fifth year. Thus, the number of deaths is greater at the beginning of the age interval than at the end, and the time lived by those dying during an age interval is less than half the age interval.

The estimation of separation factors for ages under 1 year and 1 to 4 years can be made by (a) using detailed information about the age of deceased persons; and (b) using established relationships between the level of infant mortality and the separation factors.

129
Calculation of separation factors from statistics

Separation factor for age under 1 year. It is recommended to calculate the separation factors only if there is detailed information concerning the age of the deceased infant. Infant deaths should be reported by single days of age during the first week of life, by weeks of age during the first month of life, and by month thereafter. Separation factors, then, are a weighted average of the fraction of a year lived by those dying, weighted by the number of infants dying at each age. In symbols,

\[ k_0 = \frac{\sum D_i \cdot t_i}{D} \]

Where:

- \( k_0 \) is the separation factor for age under 1 year;
- \( D_i \) is the number of infants dying within certain ages during the first year of life;
- \( t_i \) is the time lived from birth to the day the infant dies, expressed as a fraction of a year; and
- \( D \) is the total number of infant deaths during the first year of life.

For instance, if infant deaths are classified as in table A-III-2.1, then the \( t_i \) factors of column (2) are calculated as follows:

1. Infants dying during the first day of life, whose age is under 1 day, would live approximately \( \frac{1}{2} \) day (under the assumption that during the first day of life those dying are evenly distributed during the day). Consequently, the fraction of a year represented by the \( \frac{1}{2} \) day lived is \( \frac{0.5}{365} \), based on a year of 365 days. This factor is weighted by the number of deaths under 1 day of age.

2. Infants dying at age 1 day (during the second day of life) have lived, on average, \( 1\frac{1}{2} \) days since birth: 1 complete day (the first one) plus half of the second day, when they die. Hence, the factor is \( \frac{1.5}{365} \).

3. With similar reasoning, infants dying during the third day, at age 2 days, have lived on average 2.5 days, and the factor is \( \frac{2.5}{365} \).

4. For infants dying at ages 3, 4, 5, and 6 days, the factors are \( \frac{3.5}{365} \), \( \frac{4.5}{365} \), \( \frac{5.5}{365} \), and \( \frac{6.5}{365} \), respectively (table A-III-2.1, column 2).
Table A-III-2.1. Calculation of the Separation Factor for Age Under One

<table>
<thead>
<tr>
<th>Age of the deceased</th>
<th>Deaths</th>
<th>Fraction of year lived</th>
<th>Deaths times fractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>-1</td>
<td>856</td>
<td>0.5/365</td>
<td>1.2</td>
</tr>
<tr>
<td>1</td>
<td>218</td>
<td>1.5/365</td>
<td>.9</td>
</tr>
<tr>
<td>2</td>
<td>138</td>
<td>2.5/365</td>
<td>.9</td>
</tr>
<tr>
<td>3</td>
<td>68</td>
<td>3.5/365</td>
<td>.7</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
<td>4.5/365</td>
<td>.5</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>5.5/365</td>
<td>.5</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>6.5/365</td>
<td>.4</td>
</tr>
<tr>
<td>Weeks</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>1</td>
<td>101</td>
<td>1.5/52</td>
<td>2.9</td>
</tr>
<tr>
<td>2</td>
<td>87</td>
<td>2.5/52</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>3.5/52</td>
<td>5.0</td>
</tr>
<tr>
<td>Months</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>1</td>
<td>240</td>
<td>1.5/12</td>
<td>30.0</td>
</tr>
<tr>
<td>2</td>
<td>220</td>
<td>2.5/12</td>
<td>45.8</td>
</tr>
<tr>
<td>3</td>
<td>207</td>
<td>3.5/12</td>
<td>60.4</td>
</tr>
<tr>
<td>4</td>
<td>178</td>
<td>4.5/12</td>
<td>66.8</td>
</tr>
<tr>
<td>5</td>
<td>165</td>
<td>5.5/12</td>
<td>75.0</td>
</tr>
<tr>
<td>6</td>
<td>152</td>
<td>6.5/12</td>
<td>82.3</td>
</tr>
<tr>
<td>7</td>
<td>132</td>
<td>7.5/12</td>
<td>82.5</td>
</tr>
<tr>
<td>8</td>
<td>119</td>
<td>8.5/12</td>
<td>84.3</td>
</tr>
<tr>
<td>9</td>
<td>102</td>
<td>9.5/12</td>
<td>80.8</td>
</tr>
<tr>
<td>10</td>
<td>89</td>
<td>10.5/12</td>
<td>77.9</td>
</tr>
<tr>
<td>11</td>
<td>84</td>
<td>11.5/12</td>
<td>80.3</td>
</tr>
<tr>
<td>Totals</td>
<td>3,334</td>
<td></td>
<td>784.1</td>
</tr>
</tbody>
</table>

Separation factor:

\[ k_0 = \frac{784.1}{3,334} = .235 \]
(5) For infants dying during the second week of life, at age 1 week, the time lived will be the whole first week, plus half of the second week (the week they die). Hence, the factor is 1.5/52, since the year has 52 weeks. Similarly, for those dying during the third and fourth weeks, the factors are 2.5/52 and 3.5/52, respectively (table A-III-2.1, column 2).

(6) For practical purposes, infants dying after the 28th day of life but before the third month of life, are considered as dying during their second month of life, at age 1 month. They have lived 1 complete month, plus \( \frac{1}{2} \) month more, since they die during the second month of life. Hence, the factor is 1.5/12, since there are 12 months in a year.

(7) For infants dying during their third, fourth, ..., twelfth month of life (at ages 2, 3, ..., 11 months, respectively), the factors are 2.5/12, 3.5/12, ..., and 11.5/12, respectively (table A-III-2.1, column 2).

Separation factors are obtained by:

(1) multiplying the number of deaths by their corresponding factors (columns 1 and 2 of table A-III-2.1, respectively), as calculated in table A-III-2.1, column 3;

(2) summing the products; and

(3) dividing the sum by the total number of infant deaths (table A-III-2.1).

Separation factors for ages 1 to 4 years. The age interval for this group has 4 years, starting at age 1. Hence, those dying at age 1, during their second year of life, lived only \( \frac{1}{4} \) year during the age interval 1 to 4 years. The factor for these deaths is 0.5. Similarly, for those dying at ages 2, 3 and 4 years, the factors are 1.5, 2.5 and 3.5, respectively (table A-III-2.2, column 2). The calculation of the separation factor for ages 1 to 4 years is the same as for age under 1 year:

(1) the number of deaths is multiplied by the factors of lived time (table A-III-2.2, columns 1 and 2);

(2) the products are summed; and

(3) the sum is divided by the total number of deaths at ages 1 to 4 years (table A-III-2.2).
Table A-III-2.2. Calculation of the Separation Factor for the Age Group 1 to 4 Years

<table>
<thead>
<tr>
<th>Age of the deceased</th>
<th>Years lived during period (1)</th>
<th>Deaths (2)</th>
<th>Deaths times (3) (1)x(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,034</td>
<td>0.5</td>
<td>1,017</td>
</tr>
<tr>
<td>2</td>
<td>1,698</td>
<td>1.5</td>
<td>2,547</td>
</tr>
<tr>
<td>3</td>
<td>1,234</td>
<td>2.5</td>
<td>3,085</td>
</tr>
<tr>
<td>4</td>
<td>1,092</td>
<td>3.5</td>
<td>3,822</td>
</tr>
<tr>
<td>Totals</td>
<td>6,058</td>
<td></td>
<td>10,471</td>
</tr>
</tbody>
</table>

Separation factor:

\[
k_1 = \frac{10,471}{6,058} = 1.728
\]

**Calculation of separation factors from observed relationships**

From existing life tables. Since separation factors are strongly related to the levels of mortality, if the population under study does not have statistics on deaths, a separation factor can be estimated from existing life tables whose level of mortality approximates that of the population under study. In this case, using the formula on page 80 of the text, the separation factor for age under 1 year can calculated as:

\[
k_0 = \frac{L_0 - l_1}{d_0}
\]

Where:

\(L_0, l_1\) and \(d_0\) represent the functions of the life table.

The separation factor for age group 1 to 4 years can be calculated as:

\[
k_1 = \frac{4L_1 - 4 \times l_5}{4d_1}
\]

Where the symbols represent the life table functions.
In general, to calculate the separation factor for any age group \( x \) to \( x+n \) years, the formula is:

\[
\frac{\frac{nL_x - n \cdot l_{x+n}}{n_d_x}}
\]

Where the symbols represent the life table functions.

**From models and correlations.** In constructing model life tables, Coale and Demeny estimated the relationship between separation factors and the level of infant mortality rates from empirical life tables. Correlating the infant mortality rates with separation factors, they provided a set of equations for estimating model separation factors. These equations are:

(1) For infant deaths:

(a) If infant mortality is equal to or greater than 100 deaths per 1,000 live births, the separation factors are:

<table>
<thead>
<tr>
<th>Region</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>0.29</td>
<td>0.31</td>
</tr>
<tr>
<td>North, South, and West</td>
<td>0.33</td>
<td>0.35</td>
</tr>
</tbody>
</table>

(b) If infant mortality is lower than 100 deaths per 1,000 live births, the separation factors are calculated as:

\[
k_x = 0.0025 + 2.875 \cdot T_x \\
k_x = 0.01 + 3.0 \cdot T_x
\]

(2) For deaths 1 to 4 years of age:

(a) If infant mortality is equal to or greater than 100 deaths per 1,000 live births, the separation factors are:

<table>
<thead>
<tr>
<th>Region</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>1.313</td>
<td>1.324</td>
</tr>
<tr>
<td>North</td>
<td>1.558</td>
<td>1.570</td>
</tr>
<tr>
<td>South</td>
<td>1.240</td>
<td>1.239</td>
</tr>
<tr>
<td>West</td>
<td>1.352</td>
<td>1.361</td>
</tr>
</tbody>
</table>

(b) If infant mortality is lower than 100 deaths per 1,000 live births, the separation factors are calculated as:
There are several possibilities for estimating infant mortality rates, and the selection of a procedure depends on the availability and quality of information on births and deaths. If all statistics were perfect, the formula on page 66 of the text would permit the researcher to estimate age-specific death rates for all ages and, from these rates, the probabilities of dying as suggested in the formula on page 77 of the text. Unfortunately, most developing countries do not have reliable statistics or such data do not exist at all. As a consequence, the rates have to be estimated by other procedures.

**Based on reliable statistics**

One of the best procedures for estimating an infant mortality rate requires excellent information on births and deaths. Particularly, data on infant deaths occurring during a year should be not only complete but also classified by year of birth, that is, whether the deceased infant was born in the year it died or in the previous year.

One of the problems in measuring infant mortality is that two calendar years of data are required to register all possible deaths of infants born during a year. For example, most of the infants who were born during 1987 and died before celebrating their first birthday, died in 1987. However, those born during the last month of the year 1987 had a chance to die during most of the year 1988. Hence, the calculation of an infant mortality rate following a cohort of annual births is affected by the mortality conditions of two years.

Nonetheless, an infant mortality rate can be calculated based on information on infants exposed to the mortality conditions of a single year. A Lexis diagram helps in understanding the rationale of the procedure (see figure A-III-2.1). According to Figure A-III-2.1, the infant mortality rate can be calculated as 1 (one) minus the product of two survival ratios:
Figure A-III-2.1. Calculation of the Infant Mortality Rate

\[ q_0 = 1 - \frac{E^t_1}{N^t} \cdot \frac{N^{t+1}}{B^t} \]

Where:

\[ N^t = B^{t-1} - D^{t-1,t-1} \]
\[ N^{t+1} = B^t - D^{t,t} \]
\[ E^t_1 = N^t - D^{t,t-1} \]

The superscript of B represents the year of occurrence (frequently coinciding with the year of registration). Infant deaths have two superscripts: the first one represents the year of death registration, and the second one represents the year of birth of the deceased infant.
Replacing the above values in the formulas of figure A-III-2.1, the following infant mortality rate formula is obtained.

\[ q_0 = 1 - \frac{b^{t-1} - d^{t-1, t-1} - d^t, t-1}{b^{t-1} - d^{t-1, t-1}} \cdot \frac{b^t - d^t, t}{b^t} \]

Where:

- \( b^t \) represents the number births during the indicated year;
- \( t \) represents the year;
- \( d^{t, t} \) represents the number of infant deaths registered in year \( t \) and born in year \( t \).

In cases where infant deaths are not tabulated by year of birth, separation factors for infant deaths can be used. In such cases, the formula for estimating infant mortality is:

\[ q_0 = 1 - \frac{b^{t-1} - (1-k^{t-1}) \cdot d^{t-1} - k \cdot d^t}{b^{t-1} - (1-k^{t-1}) \cdot d^{t-1}} \cdot \frac{b^t - (1-k^t) \cdot d^t}{b^t} \]

Where the symbols have the same meaning as in the previous formula, except that infant deaths (\( D^t \)) have only one superscript, referring to the year of registration; and the new symbol \( k \) represents the separation factor of infant deaths for the indicated year (\( t-1 \) and \( t \)).

Frequently, the infant mortality rate is estimated by dividing the registered number of infant deaths by the registered number of births during the same year. Although this procedure is accepted for some developing countries, it is not fully satisfactory because (a) the data used are not perfectly comparable; and (b) the data may contain errors. The infant mortality rate is the proportion of births not surviving the first year of life. Hence, if infant deaths and births registered during the same year are used to obtain the infant mortality rate, then (a) some of the deaths pertain to births that occurred during the previous year; and (b) some of the infants born during the current year will die during the following year. For instance, those dying in the month of February at age over 3 months obviously were born before the beginning of the year and do not pertain to the cohort of infants born during the year the deaths occurred. Similarly, children born during the month of December, who die at ages over 1 month, will be registered as deaths during the following year.

Late registration of births is another factor producing some bias in estimating infant mortality by taking the ratio of infant deaths to births during the same year. Births registered during a year may not include all births occurring that year; also they may include some births that occurred during previous years. Similarly, not all infant deaths occurring during a year may be registered during the same year.
The possible biases mentioned above may be compensated by calculating the infant mortality rate based on information from three consecutive years, for instance, registered infant deaths from 1985 to 1987, divided by registered births for the same period. In symbols,

\[
q_0 = \frac{D^{t-1} + D^t + D^{t+1}}{B^{t-1} + B^t + B^{t+1}}
\]

Where:

- \(D\) and \(B\) represent registered infant deaths and births, respectively; and
- \(t\) refers to the year of registration.

Based on other sources of information

If vital statistics are not reliable or are unavailable, the estimation of infant mortality depends on the possibility of using information provided by censuses or surveys. Surveys may provide historical information about past births and their survivorship condition in a population. With such information, infant mortality can be estimated directly by determining how many infants did not survive up to their first birthday.

Another possibility, as explained in the text, is to estimate infant mortality based on information on children ever born and children surviving by age of mother.

Child Mortality at Ages 1 to 4 Years

Central death rates for ages 1 to 4 years may be calculated as suggested in the text, by dividing the number of registered deaths by the population of the same ages. These rates are transformed into probabilities of dying by using the proper separation factors as explained above.

However, central death rates for ages 1 to 4 may not reflect actuality because information on deaths and population for those ages may contain errors. If vital statistics on births and deaths are acceptable and children under age 5 years are not subject to migration, then child mortality can be estimated without using direct information on population. As in the case of infant mortality, Spiegelman (1968) suggested a procedure to estimate child mortality. It consists of deriving the population exposed to the risk of dying at ages 1 to 4 years during a given year by following cohorts of births through time and subtracting the deaths pertaining to each cohort. Based on this derived population and the number of deaths, central mortality rates are estimated.

Finally, another possibility is to estimate the probabilities of dying from age 1 to 4 years based on census information on children ever born and
children surviving by age of mother. As explained in the text, the computer program CEBCS from MORTPAK provides such estimates.

**Smoothing Death Rates**

Before transforming the age-specific death rates $M_x$ into probabilities of dying $q_x$, it is recommended to make a graph of the rates to notice if they are smooth. If they are not, a smoothing procedure is advised. Various smoothing processes were discussed in the text.

**Life Table**

Once the age-specific death rates are smoothed and the separation factors are available for ages under 1 year and 1 to 4 years (assuming 2.5 for the other age groups), a life table can be constructed by following the formulas presented in the text. However, a possible problem remains: how to close the open-ended age group, frequently 85+ in the case of developing countries. If a life table is not closed properly--the estimation of the $L_{x+}$ at the open-ended age group--the level of life expectancy at all ages will be affected. The younger the age of the open-ended age group, the higher the chance of misestimating the value of the life table population for the open-ended age group and of obtaining biased values of life expectancies for all ages. For this reason, it is advisable to estimate age-specific death rates up to the oldest age that the data permit. If statistics are reliable, the life table can be calculated up to advanced ages (around 100 years). Errors produced in closing a life table at very old ages (95 or 100 years) have a minimal impact on the life expectancies.

Based on the life table functions, the central death rate for the open-ended age group is

$$m_{x+} = \frac{l_x}{l_{x+}}$$

and from this relationship,

$$L_{x+} = l_x / m_{x+}$$

Under the assumption that the death rate for the open-ended age group calculated from vital statistics and population information is the same as the one in the life table, the $L_{x+}$ can be calculated.

The above formula can be used for closing the life table. Not only is it attractive because of its simplicity, but it also provides acceptable results for open-ended age groups starting with ages over 90 years. For closing the life table at younger ages, the procedure should be used with caution. At younger ages, the value of $l_x$ still may be large, so small errors in the mortality rate $m_{x+}$ may have a significant impact on the estimated $L_{x+}$ value and may affect the life expectancies. However, as mortality rates at younger ages are independent of the open-ended value, life table functions
other than the $T_x$ and life expectancies are not affected by an error in the $L_x+$ estimate.

Other attempts have been made to derive relationships based on correlations between the last available $l_x$ and the $L_x+$ or $T_x$. The United Nations and Coale and Demeny proposed the following relationships, respectively:

\[ L_{85+} = l_{85} \cdot \log_{10} l_{85} \]

\[ L_{80+} = (3.725 + 0.0000625 \times l_{80})l_{80} \]

Finally, for populations with acceptable data, the closing of the life table is sometimes performed by extrapolating the force of mortality using Gompertz-type functions. These functions are fitted to the life table functions at younger ages and are used to extrapolate the survivors of the life table up to ages where most of them have died. The latest model life tables of the United Nations used this procedure. They were closed assuming that the force of mortality at the older ages followed a Makeham function.

**Software**

There are several computer programs for estimating a life table. Two of them use information for all ages, assuming the population has reliable data. Others use only fragmentary data, and they are presented in another appendix. The programs recommended for cases where information is reliable are the Bureau of the Census spreadsheets LTPOPDT and LTMXQXAD (see documentation in volume II) and the United Nations program LIFTB in the MORTPAK package.

**Spreadsheet LTPOPDT**

**Description**

This spreadsheet constructs and smooths a life table using the procedure explained in the text. The calculations are performed for only one sex at a time, or for both sexes together. The life table is constructed based on information on population, deaths, infant mortality rate (if available), and separation factors (if available). It is advisable to provide the infant mortality rate if there is one. If no infant mortality rate is provided, the program calculates one (the $q_0$ value) based on the population and the number of deaths under age 1 year.

Separation factors for ages under 1 year and 1 to 4 years are optional. If they are available, they can be used. Otherwise, separation factors are automatically calculated using the Coale-Demeny formulas. For other ages (5-year age groups), the separation factors are assumed to be 2.5.
A logarithmic smoothing process is applied to the \(m_x\) function of the unsmoothed life table. Once the \(m_x\) values are smoothed, they are proportionally adjusted to reproduce the total number of deaths in the smoothed ages (starting with the age group 15 to 19 years).

The closing of the life table is calculated based on the open-ended central death rate. The user should be aware that, if the central death rate for the open-ended age group is "too low," the open-ended \(L_x\) value, and hence the life expectancy for the open-ended age group, may be "too high." Caution is advised.

**Data Required**

1. Populations for ages under 1 year, 1 to 4 years, and by 5-year age groups thereafter.
2. Deaths for the same age groups as the population.

**Optional**

1. Infant mortality rate if available from other sources of information.
2. Separation factors for ages under 1 year and 1 to 4 years.

**Spreadsheet LTMXQXAD**

**Description**

This spreadsheet constructs a life table from age-specific death rates or from the probabilities of dying between two specific ages. The rates must refer to ages under 1 year, 1 to 4 years, and 5-year age groups thereafter. The life table is constructed as described in the text, using central death rates or probabilities of dying and separation factors. If probabilities of dying are given as input, the closing of the life table requires the life expectancy at birth for the open-ended age group. (If a life expectancy is not available, an appropriate level from a model life table is an acceptable estimate.) In cases where age-specific mortality rates are used, the user should be aware of the problem of closing the life table with the mortality rate for an open-ended age group. If such a rate is unreasonably low, the resulting life expectancies for the older ages may be unreasonably high.

The spreadsheet allows the user to incorporate adjustment factors for correcting the age-specific death rates or the probabilities of dying, if desired. These factors are useful if it is known that the mortality rates overrepresent or underrepresent the actual levels of mortality. If the adjustment factors are left to remain as 1, the life table is constructed without modifying the input data.
Empirical separation factors for ages under 1 year and 1 to 4 years are optional. If they are available, they can be used. Otherwise, separation factors are automatically calculated using the Coale-Demeny formulas. For other ages (5-year age groups), separation factors are assumed to be 2.5.

The calculations are performed for only one sex at a time, or for both sexes together.

Data Required

(1) Age-specific death rates or probabilities of dying between exact ages, for ages under 1 year, 1 to 4 years, and 5-year age groups thereafter.

(2) Life expectancy for the open-ended age group, if probabilities of dying are used as input in (1). If age-specific death rates are used, the life expectancy is not required.

Optional

(1) Separation factors for ages under 1 year and 1 to 4 years.
Appendix III-3

Selecting a Model Life Table and Life Expectancy at Birth from Limited Data

There is a relationship among the population age and sex structure, the age-specific death rates, and the crude death rate. Given an age and sex structure and a set of age-specific death rates, only one crude death rate can be calculated. Based on such consistencies, if there is information on the age and sex distribution of the population and a crude death rate is available, it would be possible to find a model life table consistent with the population and the rate.

Four spreadsheets were developed at the Bureau of the Census; they select a pattern of mortality (central death rates or age-specific death rates) that, when applied to the population, reproduce the crude death rate or a given number of deaths. The four spreadsheets are ADJMX, LTNTH, LTSTH, and LTWST. They are described below.

Spreadsheet ADJMX

Description

A pattern of mortality by age (age-specific death rates or central death rates from an empirical or existing life table) is applied to the sex and age structure of a population. The calculated total number of deaths is compared with the total number of deaths pertaining to the population. If there is a difference, the pattern of mortality is adjusted proportionally. When the adjusted pattern is applied to the population age structure, the resulting number of deaths is the same as that pertaining to the population. Then, the program generates life tables for each sex (or for both sexes if information by sex is not available).

The program matches the number of deaths for each sex separately. However, if the total number of deaths is available only for both sexes combined, the program distributes the deaths by sex in proportion to the calculated deaths using the information on population and age-specific death rates.

Data Required

1. Age-specific death rates by sex. Age groups should be under 1 year, 1 to 4 years, and 5-year age groups thereafter.

2. Population by sex and age. Age groups same as in (1).
(3) Total number of deaths for each sex or for both sexes combined. Use the figure for both sexes only if the number of deaths is not available by sex.

Software

See documentation of ADJMX in volume II.

Spreadsheets LTNTNTH, LTSTH, and LTWST

Description

These spreadsheets estimate a life expectancy at birth from crude death rates, the age structure of the population, and a pattern of mortality. They can be used for populations of small areas where deaths are not tabulated by age. In addition to the life expectancy at birth, estimates of the crude birth rate and the infant mortality rate are provided. The spreadsheet prints the model life tables used in the estimation process.

Data Required

1. Population by sex and age. Age groups should be under 1 year, 1 to 4 years, and 5-year age groups thereafter.
2. The population's crude death rates for each sex or both sexes combined.

Procedure

These spreadsheets follow a procedure similar to the one described for the spreadsheet ADJMX. However, instead of requiring a pattern of mortality, these spreadsheets use model patterns. The models used are three of those developed by Coale and Demeny: North, South, and West models in the spreadsheets LTNTNTH, LTSTH, and LTWST, respectively.

Given the male and female age structure of the population and the crude death rates for each sex or for the whole population, each program searches for a set of model central death rates that, when applied to the population, reproduces the given crude death rates. The program calculates and provides the model life table for each sex that reproduces the crude death rates and the life expectancy at birth for each sex.

The output from these spreadsheets includes some indices for evaluating the input data. In addition to the output mentioned above, the infant mortality rate, crude birth rate, and rate of natural
increase are provided. The infant mortality rates correspond to the selected model life tables for each sex. The birth rate is estimated as follows:

The number of births during the year prior to the census:

\[ B' = P_0 \cdot \frac{l_0}{L_0} \]

Where:

- \( P_0 \) is the population under age 1 year, and
- \( l_0 \) and \( L_0 \) are the functions of the life table for age 0.

These births are centered a half year before the census. In order to have them centered on the census date, they are multiplied by the following factor:

\[ B = B' \exp\left(\frac{r'}{2}\right) \]

Where:

- \( B \) is the estimate of the annual birth centered at census date, and
- \( r' \) is an estimate of the rate of natural increase.

This estimate of the rate of natural increase is calculated as:

\[ r' = b' - d \]

Where:

- \( d \) is the crude death rate calculated from the deaths obtained by multiplying the population times the central death rates of the selected life table and the total population, and
- \( b' \) is an estimate of the birth rate obtained as \( b' = B'/P \) where \( P \) is the total population.

This formula for estimating the crude birth rate depends on the population under age 1 year. If this population is underenumerated, the birth rate will be underestimated. Thus, if this spreadsheet provides a lower than expected crude birth rate for countries where an approximate level of the crude birth rate is known, it is because the population under age 1 year should be larger. Although the birth rate is affected also by the level of mortality, it is much more sensitive to the population under age 1 year.
Software

See documentation of LTNTH, LTSTH, and LTWST in volume II.
Appendix III-4

Stable Population

Stable population theory was derived during the first decades of this century based on observations of microbiological populations (Lotka, 1934). Later, it was expanded to human populations under the assumption that they have constant mortality and fertility and are not exposed to migration. Since then, several books and articles have expanded the application of stable population theory to demographic analysis.

Stable population theory has proved that if, beginning at any moment in time, a population with a certain age structure sustains constant levels of mortality and fertility for a long period of time, without migration, it will develop a new age structure. This new age structure will be independent of the old one and will remain unchanged as long as mortality and fertility remain constant. In other words, the age structure depends only on the levels of mortality and fertility, provided that they remain constant for a long period of time.

Since a stable population has constant birth and death rates, its growth rate is also constant. In this theory, the three basic rates (growth rate, crude death rate, and crude birth rate) are called intrinsic rates.

The calculation of a stable population for females should start with the conditions mentioned above, namely that mortality and fertility are constant. Hence, the first step is to calculate the intrinsic growth rate by solving the following equation (in the discrete field) for a value of \( r \), the intrinsic growth rate.

\[
1 - \sum_{x=15}^{50} n f_x n L_x e^{-r(x+n/2)}
\]

Where:
- \( x \) represents age;
- \( n \) represents the years of the age interval;
- \( f \) is the age-specific fertility rate for female births;
- \( L \) is the female life table function; and
- \( r \) is the intrinsic growth rate.

The equation is solved by an iterative process, substituting values of \( r \) until both sides of the equation are equal.

The next step is to calculate the intrinsic birth rate, \( b \).
The last value of the L function is the open-ended age group. In this case, the exponent \( x \) is the beginning age of the open-ended age group, and \( \frac{n}{2} \) is the life expectancy at age \( x \).

If \( C(x, x+n) \) represents the proportion of the stable population in the age group \( x \) to \( x+n \) years, then, the proportional age distribution of the stable population is calculated as follows:

\[
C(x, x+n) = b_n L_x e^{-r(x+n/2)}
\]

For the open-ended age group, the value of \( n/2 \) is replaced by the life expectancy at the beginning age of the open-ended age group.

The intrinsic death rate is calculated as the difference between the intrinsic birth rate and the intrinsic growth rate.

For males, stable population theory would require male fertility rates to follow the same procedure. However, a different intrinsic growth rate would be obtained. Therefore, if females and males have different intrinsic growth rates, the age structure of the population for both sexes combined would not remain constant. The age structure would tend to approach the age structure of the sex with the higher intrinsic growth rate, and hence it would not be a stable population.

To avoid this problem, the practical solution adopted is to calculate the male stable population by using the female intrinsic growth rate. The formulas are the same as for females.

**Software**

The Bureau of the Census has developed a spreadsheet to calculate a stable population based on fertility and mortality rates, following the procedure explained above. The name of the spreadsheet is SP, and its documentation is presented in volume II.
Appendix III-5

Preston-Coale Technique

Description

This technique estimates the completeness of reporting of deaths in relation to information on population (United Nations, 1983). It compares a distribution of deaths with a corresponding population, by 5-year age groups. Formulas based on stable population theory are applied to the available information on deaths to obtain estimated populations at certain ages. These population estimates are compared with the actual populations, and the differences are attributed to incompleteness of the information on deaths.

Data Required

1. Population by 5-year age groups.
2. Deaths by 5-year age groups.
3. A population growth rate.

Assumptions

1. The population has stable characteristics: mortality and fertility were constant during the past, and there was no migration.
2. Underenumeration of the population is the same in all age groups.
3. Underenumeration of deaths is the same in all age groups.

Procedure

Complete the following steps to estimate the underenumeration of deaths:

1. Calculate the required population growth rate based on population information from two censuses. The authors recommend (a) to use the population over age 10 or 15 years to estimate the growth rate; or (b) to calculate growth rates for the population over ages 10, 15, ..., 60 years and then to select the median growth rate.

2. An optional step is to adjust the population to the midpoint of the year to which the deaths refer. As the purpose is to evaluate the reporting of deaths in relation to the population to obtain age-specific mortality rates, this step usually is not needed.
(3) Based on the number of deaths and growth rates, estimate the population for all ages except the open-ended age group as:

\[ EP_x = EP_{x+5} \cdot \exp(5r) + D_{x,x+4} \cdot \exp(2.5r) \]

Where:

- \( EP \) is the estimated population at exact age \( x \);
- \( r \) is the growth rate; and
- \( D \) represents the number deaths.

(4) For the open-ended age group \( x+ \), estimate the population at age \( x \) as:

\[ EP_x = D_{x+} \cdot \exp[r \cdot z_x] \]

Where the factor of the exponent \( z_x \) was estimated by the authors using correlation coefficients obtained from stable population based on the mortality from Coale-Demeny model life tables and relating the factor \( z_x \) to the growth rate and to the exponential of the ratio of deaths for ages 45+ and 10+.

(5) Use the estimated populations at exact ages \( EP_x \) to estimate the population in 5-year age groups as:

\[ EP_{x,x+4} = 2.5 \cdot [EP_x + EP_{x+5}] \]

(6) Calculate the completeness of death reporting by dividing the estimated total population by the actual total population in the same ages:

\[ C = EP/P \]

Where \( EP \) and \( P \) represent the total estimated and actual populations, respectively. These totals usually pertain to ages 5 years and over.

(7) To adjust the number of deaths for underreporting, divide the number of registered deaths by the factor \( C \).

Suggestions

The ratios mentioned above in step (6) should be calculated for all age groups. When these ratios for different ages resemble a horizontal straight line, they support the results of the technique. The departure of these ratios from a horizontal straight line indicates either that the population is not stable or quasi-stable, or that there are serious errors in the data (such as differential underenumeration of population and registration of deaths by age). In this case, the results must be interpreted with caution.
Advantages

The technique provides an evaluation of the information on deaths based on only one census of population.

Limitations

The results will be affected by the following conditions:

(1) Different degrees of completeness in the population and death information by age.
(2) Strong age misreporting of both population and deaths.
(3) Lack of stable or quasi-stable condition of the population.
(4) Lack of a good estimate of the population growth rate.

Software

The Bureau of the Census has developed a spreadsheet to apply this technique to a distribution of deaths and population. The spreadsheet is called PRECOA and its documentation is given in volume II.
Appendix III-6

Growth Balance Mortality Technique

Description

This technique estimates the completeness of reporting of deaths over age 5 years in relation to information on population (Brass, 1975). It compares the distribution of deaths in relation to the distribution of population, both by age. Deaths usually pertain to a period of 1 year and, if possible, data on both deaths and population should refer to the same year.

Data Required

(1) Deaths by age, preferably by 5-year age groups. Deaths can be for both sexes combined or for each sex individually.

(2) Population with the same age and sex breakdown as deaths. If possible, the time reference should be midyear of the year to which the deaths pertain.

Assumptions

(1) The population has stable characteristics: mortality and fertility were constant during the past, and there was no migration.

(2) Completeness of death registration is the same for all age groups over age 5 or 10 years.

(3) There is no age misreporting of the population or of deaths.

Procedure

Accepting the above assumptions, estimate the completeness of reporting of deaths over age 5 or 10 years as follows:

(1) Calculate the population at exact ages ending in digits 0 and 5, for ages 5 years and above, as follows:

Sum the population for two consecutive age groups and divide the sum by the number of years spanning the two age groups (for example, for two 5-year age groups, divide by 10). Other interpolation methods can also be used.

(2) Calculate the cumulated population for ages x and over by cumulating the population reported in each age group. For 5-year age groups, do this for ages 5+, 10+, 15+, and so forth.
(3) Calculate the cumulated number of deaths for the same ages as was done for the population.

(4) Calculate the ratios of the cumulated deaths and population, by age \((\text{partial death rates})\), as well as the ratios of the population at exact age \(x\) to the cumulated population for the same age \(x+\) \((\text{partial birth rates})\).

(5) Plot these ratios on a graph, as follows:

\[
\begin{align*}
\text{On Y-axis:} & \quad \text{partial birth rates} \\
\text{On X-axis:} & \quad \text{partial death rates}
\end{align*}
\]

If the population meets the assumptions of the technique, the points of the graph should lie on a straight line. Deviations from linearity are expected because actual populations are not stable and because the data usually have errors. Results of the technique should be accepted only if the points lie close to a straight line, particularly those points pertaining to ages 10 to 55 years. The points may fail to lie on a straight line as a consequence of:

(a) age misreporting;

(b) non-stability of the population; or

(c) differential completeness in reporting of deaths by age.

(6) Select the best fitting line, as follows:

(a) Examine the graph. If a large proportion of the points lie on or near a straight line, fit a straight line to the points. Choose at least 9 points, if possible.

(b) To fit the straight line, follow either one of two methods:

(i) Use least squares.

(ii) Compute averages of selected points to derive two average points and use the two to calculate the straight line equation. For instance, if there are 10 acceptable points, make two groups of 5 points each and take the average of the abscissas and ordinates of the points. The average coordinates represent the average point of the groups. If there are 9 points, group the first 3 (or 4) points and the last 3 (or 4) points, and take the appropriate averages.

(7) Calculate the completeness of reporting of deaths by estimating the slope of the line, as follows:
(a) If the method used was least squares, take the slope and follow the instructions in step (c), below.

(b) If the method used was the average, calculate the slope of the line passing through the two average points, and then follow the instructions in step (c), below.

(c) Obtain the degree of completeness of reporting of deaths by taking the reciprocal of the slope (one divided by the slope).

(8) Adjust the number of deaths, as follows:

Multiply the number of deaths by the estimated adjustment factor for deaths from step 7 (c), above. This adjusted number of deaths should be used to calculate the age-specific mortality rates.

**Advantages**

This technique provides information on the quality of the data and permits an adjustment in cases where the population meets the assumptions made in developing the method.

**Limitations**

(1) A rapid change in mortality may produce a bias in the estimation of the completeness of reporting of deaths, but slow changes in mortality over a long period of time will have only a small impact on the completeness estimate.

(2) Recent changes in fertility will cause the points that include the younger ages not to be in line with the other points. However, the results still can be used if they are based on the ages which were not affected by the change in fertility.

(3) Migration will have an effect on the results if the age structure of the migrants differs from that of the population.

(4) Age misreporting and differential completeness of reporting of the population by age may have the largest impact on the estimation of the factor for adjusting the number of deaths.

(5) The estimation of completeness refers only to deaths at ages 5 years and above. Infant and child deaths are not evaluated by this technique.
Software

There is a microcomputer spreadsheet that may be used to make all the calculations to estimate the completeness of reporting of deaths. The program is called GRBAL, and its documentation is presented in volume II.
Appendix III-7

Bennett-Horiuchi Mortality Technique

Description

This technique estimates the completeness of death registration above a certain age \( x \) during an intercensal period, based on population distributions from two consecutive censuses (Bennett and Horiuchi, 1981). Age \( x \) is the age above which the degree of completeness of death registration can be assumed to be uniform, usually taken to be age 5 years. The technique also provides a set of adjusted death rates by age, as well as estimated life expectancies for ages 5 years and above during the intercensal period.

Data Required

(1) Population age distributions from two consecutive censuses, by 5-year age groups.

(2) Number of registered deaths during the intercensal period, by 5-year age groups.

Assumptions

(1) The population was not exposed to migration during the intercensal period.

(2) Both censuses have the same degree of completeness.

(3) Age misreporting occurs only after age 50 years.

(4) Degree of completeness of death registration is uniform above age 5 years.

Procedure

The method uses the number of registered deaths and population growth rates for each age group (same ages in both censuses), to estimate the expected population at certain ages. A comparison of the expected population with the population enumerated in the censuses provides the degree of completeness of death registration.

Advantages

(1) The technique does not assume that the population is stable. This is an advantage over similar techniques that require such an assumption.

157
(2) In addition to estimating completeness of death registration, the results may be used to evaluate the base information used. If the estimated degree of completeness at different ages is similar, the base information may be considered consistent. If the estimates of completeness differ considerably from age to age, then the base information may contain errors or may not meet the assumptions of the technique. If this is the case, the results should be used with caution.

Limitations

(1) Migration has an effect on the estimation of completeness of death registration. If there has been immigration, the result will indicate a better degree of completeness than is actually true. Outmigration will mislead the user to believe that death registration is less complete than it actually is.

(2) The technique is sensitive to different degrees of enumeration in the two consecutive censuses:

(a) Relative underenumeration in the first census (or overenumeration in the second) would raise the estimated degree of completeness of death registration.

(b) Relative overenumeration in the first census (or underenumeration in the second) would reduce the estimated degree of completeness of death registration.

(3) The technique does not evaluate completeness of registration of deaths under age 5 years. Infant and child deaths usually have a higher degree of underregistration than deaths at other ages.

Software

The computer program BENHR in the United Nations MORTPAK will make the calculations to estimate the completeness of death registration.
Appendix III-8

Integrated Technique

Description

This technique, developed by Preston (1983), estimates consistent demographic parameters. Based on the age structure from two consecutive censuses, the method estimates the intercensal level of mortality, the crude birth rate, and an age distribution of the population for the intercensal period. By using logits, the technique combines the Brass general mortality pattern with a generalized stable population equation developed by Preston and Coale in 1982. Because the method "integrates" the Brass mortality model with the generalized stable population, it is called the integrated technique. The level of mortality is estimated using life expectancies for age 5 years and above.

Data Required

(1) The age distribution of the population from two consecutive censuses, by 5-year age groups.

(2) A model or empirical pattern of mortality. The pattern does not have to represent the level of mortality of the population, although it is advised to choose an "expected" level. The author advises also to use an estimate of child mortality.

Assumptions

(1) The population was not exposed to migration during the intercensal period.

(2) Both censuses have the same degree of completeness of enumeration.

Procedure

The method estimates intercensal growth rates for each 5-year age group. These rates are used in the generalized stable population equation. By using the one-parameter logit system of the Brass mortality model and the generalized stable population equation, a linear relationship is derived, relating mortality and fertility. The reciprocal of the intercept of a line fitted to the data is the crude birth rate. The slope of this line represents an adjustment factor for the pattern of mortality selected. After adjusting the pattern, life expectancies for ages 5 years and above can be calculated. Consistency of the original data is determined by examining the position of the points used for fitting the straight line. The closer the points are to a straight line, the more consistent the data are with the assumptions of the technique.
**Advantages**

(1) The technique does not assume that the population is stable. It only requires a closed population (no migration).

(2) The technique provides an estimate of mortality, a birth rate, and an age structure of the population.

**Limitations**

(1) Intercensal migration has a significant effect on the mortality level estimated (but less effect on the crude birth rate).

(2) Age misreporting has an impact on the estimated mortality level and crude birth rate.

(3) The method is sensitive to differential completeness of enumeration of the two censuses. It is more sensitive to differential completeness than the method developed by Bennett and Horiuchi.

**Software**

There is a computer program that makes all the calculations to apply this technique. It is the PRESTO program in the United Nations MORTPAK package.
Appendix III-9

Preston-Bennett Mortality Technique

Description

This method estimates the level of mortality for ages 5 years and above during an intercensal period (Preston and Bennett, 1983). Life expectancies for ages 5 years and above are obtained based on age distributions of the population from two consecutive censuses.

Data Required

The 5-year age distribution of the population from two consecutive censuses.

Assumptions

(1) Both population censuses have the same degree of completeness of enumeration.

(2) Completeness of enumeration is the same for all ages over 5 years.

(3) Age misreporting is the same in both censuses.

(4) There was no migration during the intercensal period.

Procedure

Based on the population distributions by 5-year age groups from the two censuses, intercensal growth rates for each age group are calculated. These growth rates are used, together with the information on population, to estimate the population equivalent to the life table \( L_x \) values. The population at exact ages is also estimated, representing the \( l_x \) values of the life table. With estimates of these two functions, life expectancies at ages 5 years and above are calculated.

Advantages

(1) Only the age structures of two censuses are required.

(2) The technique does not assume the population is stable, nor does it use model life tables.
Limitations

(1) The levels of mortality are affected by differential completeness of enumeration in the two censuses.

(2) Differential patterns of age misreporting in the two censuses can produce undesirable fluctuations in the estimated life expectancies at different ages.

(3) Intercensal migration affects the estimated levels of mortality.

Software

The Bureau of the Census has developed a computer spreadsheet that makes the calculations of this technique. It is called PREBEN, and its documentation is presented in volume II.
Appendix III-10

Techniques for Indirect Estimation of Infant and Child Mortality

Description

There are several techniques for estimating infant and child mortality (Brass, 1975; Coale and Trussell, 1977; Feeney, 1976; Palloni and Heligman, 1985; Sullivan, 1972; United Nations, 1983). All of them use data on the average number of children ever born alive, by age of mother, and the average number of children surviving at the time of the census or survey, by age of mother. This information is used to estimate the proportion of children who have died. Based on this proportion and assuming certain fertility and mortality patterns, the probabilities of dying between birth and certain ages are estimated.

Data Required

1. Number of children ever born alive, classified by 5-year age groups of mother (15-19 years to 45-49 years), collected in a census or survey.

2. Number of children who have survived up to the time of the same census or survey, classified by 5-year age groups of mother.

3. Number of women, by 5-year age groups, from the same census or survey.

Assumptions

1. The risk of a child dying is a function only of the age of the child and not of other factors (for example, mother's age or child's birth order).

2. Information on children ever born and on children surviving, by age of mother, are equally well reported.

3. Fertility patterns and levels have remained constant for at least 15 to 20 years before the census or survey.

4. The age pattern of mortality is the same as in a selected model life table.

5. The Brass, Sullivan, and Trussell methods originally assumed that mortality has been constant up to the present. However, Feeney (1976) explicitly built in an assumption of linearly changing mortality. Coale and Trussell (1977) then applied a similar method to allow dating of the estimates based on the Coale-Demeny model life tables. Palloni and Heligman (1985) developed similar

Procedure

William Brass noticed that the proportions dead of children ever born, classified by age of mother, were close to the probability of dying between birth and certain ages, and that the differences were primarily a function of the pattern of fertility. He therefore developed a set of multipliers to convert the proportions dead to the life table \( xq_0 \) values, the probability of dying between birth and age \( x \):

\[
xq_0 = k_i \cdot D_i
\]

Where:
- \( xq_0 \) is the probability of dying between birth and age \( x \);
- \( k_i \) is a multiplier;
- \( D_i \) is the proportion of children dead; and
- \( i \) is the age group of the mother.

The \( k_i \) multipliers vary as a function of the fertility pattern in the population being studied. The multipliers are usually estimated based on ratios of successive average parities, \( P(i)/P(i+1) \), where \( i \) is the index of the age group (1 = 15-19, ... 7 = 45-49). The data for mothers at selected age groups provide the following information:

<table>
<thead>
<tr>
<th>Index</th>
<th>Age of mother</th>
<th>Derived values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15-19</td>
<td>( lq_0 ) probability of dying between birth and age 1</td>
</tr>
<tr>
<td>2</td>
<td>20-24</td>
<td>( 2q_0 ) probability of dying between birth and age 2</td>
</tr>
<tr>
<td>3</td>
<td>25-29</td>
<td>( 3q_0 ) probability of dying between birth and age 3</td>
</tr>
<tr>
<td>4</td>
<td>30-34</td>
<td>( 4q_0 ) probability of dying between birth and age 4</td>
</tr>
<tr>
<td>5</td>
<td>35-39</td>
<td>( 5q_0 ) probability of dying between birth and age 5</td>
</tr>
<tr>
<td>6</td>
<td>40-44</td>
<td>( 6q_0 ) probability of dying between birth and age 6</td>
</tr>
<tr>
<td>7</td>
<td>45-49</td>
<td>( 7q_0 ) probability of dying between birth and age 7</td>
</tr>
</tbody>
</table>

These data can be used to estimate the infant mortality rate and the life expectancy at birth by matching the derived mortality measures with a model life table. The procedure assumes that the chosen model life table accurately reflects the pattern of mortality of the population being analyzed.

The childhood mortality estimates resulting from these techniques pertain not to the year of data collection but to a period running from some
date in the past up to the date of the census or survey. A reference date for each set of childhood mortality figures can be estimated by making certain assumptions. Such time reference estimates are included in the CEBCS program of the United Nations MORTPAK package.

Another possibility for estimating infant and child mortality is to use data on duration of marriage; the procedure for developing the technique is the same.

**Advantages**

1. Only a small amount of data is needed to apply these techniques.
2. A trend in mortality may be observed from the results.
3. Since the estimates are based on data on births for several years, they may be less affected by sampling error than methods based on data for only 1 year.
4. Mortality for social groups can be analyzed if proper tabulations are made.

**Limitations**

1. The estimate of infant mortality based on information provided by women ages 15 to 19 years should be interpreted with caution. Usually children born to these young women have higher mortality than those born to older women. This appears to be true not only for biological reasons, but mainly because the socioeconomic characteristics of the youngest mothers may be related to higher infant mortality. The estimates for the 15 to 19 year age group may also be more severely affected by sampling variation (since the youngest women have fewer births) and more sensitive to deviations of the actual pattern of fertility from the pattern implied by the model.
2. If the pattern of fertility implied by the data has not been constant (due either to a change in the level of fertility or to a change in the pattern), the results may be affected.
3. Poor quality of the data will produce results of uncertain reliability:
   
   (a) If the estimates of child mortality fluctuate erratically, this is more likely due to data problems than to actual mortality changes.
   
   (b) If the estimates imply that mortality was lower in the past than the present, then the data to which the technique was applied should be carefully evaluated for possible errors.
In particular, older women may be more likely than younger women to underreport the number of children who died, especially those who died when the women were very young.

(c) Age misreporting of mothers can affect the results, since the children's length of exposure to the risk of dying is inferred from their mothers' ages.

<table>
<thead>
<tr>
<th>Summary of Software</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author and program</strong></td>
</tr>
<tr>
<td>Brass</td>
</tr>
<tr>
<td>CEBCS-MORTPAK</td>
</tr>
<tr>
<td>Trussell</td>
</tr>
<tr>
<td>Palloni-Heligman</td>
</tr>
<tr>
<td>CEBCS-MORTPAK</td>
</tr>
</tbody>
</table>

166
Appendix III-11

Johnson Technique

Description

This technique estimates infant mortality rates based on information on the number of children born during a year and the number still alive at the end of the year (U.S. Bureau of the Census, 1982). Under these circumstances, some births have been exposed to mortality for almost a full year, while others (those born close to the end of the year) have been exposed for only a few days or minutes. Based on life table concepts, and noting the similarity between information on population and the $l_0$ and $L_0$ life table functions, this technique estimates infant mortality rates using separation factors of deaths.

Data Required

(1) Census or survey population under age 1 year.

(2) Births during the 12-month period prior to the census or survey.

Assumptions

All births during the 12 months prior to the census or survey date and all the population under age 1 year are reported. This condition may not be required if the proportion of underreporting is the same for births and population and if the proportion dead among reported births is the same as among unreported births.

Procedure

The technique is based on the following life table relationship:

$$d_0 = (l_0 - L_0) / (1 - k_0)$$

Where:

- $d_0$, $l_0$, and $L_0$ are the life table functions (see section of text on Life Table in chapter III); and
- $k_0$ is the separation factor of deaths for age under 1 year.

The births and population for age under 1 year reported in the census or survey are similar to the $l_0$ and $L_0$, and the separation factors can be estimated from the Coale-Demeny formulas. Thus, the number of deaths pertaining to a year, and hence, the infant mortality rate, can be estimated.
Advantages

(1) The technique does not rely on any model pattern of mortality.

(2) The technique provides estimates of infant mortality not only by groups of women as other techniques do, but also by age of mothers.

Limitations

The technique will be affected if:

(1) The reference period for reporting births is not 12 months.

(2) Some of the births of children who die had not been reported.

Software

A spreadsheet developed by the Bureau of the Census performs the calculations of this technique. It is called BTHSRV, and its documentation is presented in volume II.
Appendix III-12

Orphanhood and Widowhood Techniques

Orphanhood Technique

Description

The orphanhood technique estimates the level of adult female mortality based on the proportion of persons declaring that their mother is still alive (United Nations, 1983). This proportion represents a group of mothers who have survived a certain period of time. Using simulation models that take into account fertility and mortality models, the proportion of persons with mothers still alive can be closely related to a survival ratio from age 25 years to some older age.

Data Required

1. Proportion of population age 15 to 49 years whose mother is still alive, by 5-year age groups.
2. Mean age of mothers during a specified period, usually 1 year.

Assumptions

1. There is no age misreporting.
2. Mortality and fertility have not changed during the past.
3. Migrants have the same age distribution and the same characteristics concerning orphanhood as the whole population. In other words, migrants represent a sample of the population.

Procedure

Several relationships have been developed to transform the proportion of persons who are not orphans into the probability of surviving from age 25 years to an older age. The first relationship was derived by Brass:

\[
\frac{1_{25+x}}{1_{25}} = W_x S_{x-5,x-1} + [1-W_x] S_{x,x+4}
\]

Where:

\( x \) represents age, usually starting at 15 years;
$S_{x,x+4}$ is the proportion not orphaned in the 5-year age group; and

$W_x$ is a weighting factor for transforming the proportions of non-orphaned population into probabilities of surviving. The $W_x$ factors were estimated using the Brass general standard model of mortality and a third degree polynomial representing the fertility patterns.

Later, Hill and Trussell used other mortality and fertility patterns (Coale-Demeny model life tables and Coale-Trussell fertility schedules) to simulate the relationship and find correlation coefficients relating the probability of surviving to the proportion of persons not orphaned:

$$l_{25+x} / l_{25} = a(x) + b(x) M + c(x) S_{x-5,x-1}$$

Where:

$a$, $b$, and $c$ are correlation coefficients; and

$M$ is the mean age of mothers during a year.

These coefficients, as well as the previous ones, are available in a United Nations publication (United Nations, 1983).

**Advantages**

Offers the possibility of estimating some indices of adult mortality.

**Limitations**

(1) Mortality and fertility may not have been constant during the past.

(2) Migrants may constitute a special group of persons with characteristics concerning orphanhood different from those of persons who did not migrate.

(3) Orphanhood information may not be accurate in countries with high internal migration and poor communication systems.

(4) If the population has a mortality pattern significantly different from those of model life tables, the estimates will be affected.
Software

The computer program ORPHAN in the United Nations MORTPAK package estimates adult mortality using this technique with the correlation coefficients calculated by Hill and Trussell.

Widowhood Technique

Description

Like the previous technique, this one estimates adult mortality of a population (United Nations, 1983). The information used in this case is whether or not a person’s first spouse is still alive. Under certain assumptions concerning the age of spouses, the proportion of widows or widowers indicates the survivorship chances a deceased spouse has had. Based on these proportions and certain coefficients, an estimate of adult mortality can be obtained.

Data Required

(1) The singulate mean age at marriage for males and females.

(2) The proportion of males or females whose first spouse is still alive.

Assumptions

(1) Mortality and age at first marriage have remained constant during the past.

(2) Migrants' characteristics concerning marital status are the same as those of nonmigrants at each age.

Procedure

Hill and Trussell obtained correlation coefficients to apply to the singulate mean at marriage and the proportion of widows or widowers to derive the probability of surviving between two adult ages. In symbols, for males:

\[
\frac{m_{1x}}{m_{120}} = a(x) + b(x) (\hat{e}_M) + c(x) (\hat{m}_M) + d(x) [m_{NWx-5,x-1}]
\]

and for females:

\[
\frac{f_{1x}}{f_{120}} = a(x) + b(x) (\hat{e}_M) + c(x) (\hat{m}_M) + d(x) [f_{NWx,x+4}]
\]
Where:

\[ x \]
indicates age;

\[ a, b, c, \text{ and } d \]
are correlation coefficients;

\[ \bar{M} \text{ and } \bar{M} \]
represent the singulate mean age at marriage for females and males, respectively; and

\[ \bar{NW} \text{ and } \bar{NW} \]
are the proportions not widowed for females and males, respectively.

The above formula requires information on the singulate mean age at marriage. Hajnal developed a procedure to estimate the mean age at first marriage based on information on ever-married persons from censuses or surveys (Hajnal, 1953). The formula is:

\[ M = \frac{(30 + 10xPS - 50xS)}{(2 - S)} \]

Where:

\[ PS \]
is the sum of the proportion single in 5-year age groups from age 15 to 19 years to age 45 to 49 years; and

\[ S \]
is the sum of the proportions of single persons in the age groups 45 to 49 years and 50 to 54 years.

**Advantages**

Provides an estimate of adult mortality for ages above 20 years.

**Limitations**

1. The technique requires a special question on the census questionnaire. The information required is the survivorship of the first spouse, which should not be mistakenly equated with the proportion of the population that is not widowed at the time of the census. Divorced persons may not know the survival status of their first spouse.

2. Age misreporting will usually produce biased results.

3. The technique requires a calculation of the singulate mean age at marriage based on the proportion of population ever married. If separated couples or divorced people report themselves as single, the singulate mean age at marriage, and hence the estimated mortality level, will be affected.
(4) The technique assumes that the population has a particular pattern of mortality, provided by model life tables. If the population has a different pattern of mortality, the estimates may be biased.

(5) The technique assumes that marriage patterns have not changed during the past. Changes in marriage patterns will affect the estimates.

**Software**

The computer program WIDOW in the United Nations MORTPAK package can be used to apply this technique.
Appendix III-13

Model Life Tables

As mentioned in the text, there are two frequently used sets of model life tables: the Coale-Demeny regional model life tables and the 1982 United Nations model life tables for developing countries. Both sets are used for various theoretical and estimation purposes. In demographic estimation, fragmentary information on mortality is compared with the models to estimate infant mortality and other specific parameters. Model mortality patterns are used in reconstructing life tables. For projection purposes, model life tables are used to estimate future patterns of mortality corresponding to specific levels of life expectancies at birth.

Other model life tables are available, but they are used less frequently. They include the original United Nations, the Ledermann, the Organization for Economic Cooperation and Development (OECD), and the logit models.

The more extensive use of the 1982 United Nations and the Coale-Demeny model life tables is attributable to their easy accessibility through computer programs (see Software, below). In addition to the MATCH program, other programs in the MORTPAK package match mortality estimates at certain ages with values in a model life table and then select the corresponding values of infant mortality and life expectancy at birth from the model.

Software

The United Nations computer package MORTPAK has a program MATCH, which generates any model life table developed by the United Nations or Coale-Demeny.
Appendix III-14

Smoothing with Logits

The procedures for smoothing with logits require that a pattern of mortality be available, to be used as a standard. There are three microcomputer programs to smooth a set of function values from a life table. Two of them are spreadsheets (LOGITLX AND LOGITQX) developed at the Bureau of the Census, and the third is BESTFT in the MORTPAK package of the United Nations. Documentation of the first two programs is presented in volume II.

Spreadsheets LOGITQX and LOGITLX

Procedure

Both spreadsheets use the same logit technique explained in the text. The difference between the spreadsheets is only in the life table function used in calculating the logits. While LOGITQX uses the probability of dying ($q_x$), LOGITLX uses the survivors at exact ages ($l_x$). Once the logits of the standard and the actual data are taken, a linear relationship between them is found by using the procedure of grouped mean.

The smoothing process is as follows:

1. It is assumed that the relationship between the logits of a life table function from two different life tables is linear, as

   \[ Y_1 = a + b Y_2 \]

   Where $Y_1$ and $Y_2$ are the logits of the empirical and standard life tables.

2. Constants $a$ and $b$ are calculated by the grouped mean technique.

3. With the values of $a$ and $b$, and the logits of the standard life tables, the smoothed values of the empirical life table are calculated.

The smoothed values practically never reproduce any of the empirical values. Sometimes it would be desirable to reproduce the same infant mortality rate as the one given as input (which can be reliable). For this purpose, both spreadsheets use the logit system applied to the smoothed values for calculating life tables that will reproduce the empirical infant mortality rate. Two life tables are offered by the spreadsheets: the one that is only smoothed or the one that is both smoothed and adjusted to the desired level of infant mortality rate.
The user should be aware that different life tables are estimated by using $l_x$ values or $q_x$ values in the smoothing process. This could be a disadvantage when smoothing with logits.

**Data Required**

1. A set of $nq_x$ values for LOGITQX, or a set of $l_x$ values for LOGITLX. These sets are the ones to be smoothed.

2. Another set of $nq_x$ values (for LOGITQX) or $l_x$ values (for LOGITLX) to be taken as the standard set. The sets taken as a standard could be from any life table (empirical or model). It is advisable that the pattern and level of the standard mortality be as close as possible to the actual pattern and level of mortality of the population (if the level and pattern are known).

3. (Optional). An empirical infant mortality rate.

**Comments**

Whether or not the user of these programs should choose a life table that reproduces the empirical infant mortality rate depends on the reliability of the empirical information. Sometimes the infant mortality rate is a reliable estimate, and the user may want to keep it. The usual options are:

1. Accept the smoothed life table which reproduces the desired or empirical infant mortality rate.

2. Accept the life table provided by the smoothing procedure, whatever the infant mortality rate.

3. Accept the life table values for age 1 year and above (or age 5 years and above) from option (2) and the infant mortality rate (and mortality for ages 1 to 4 years) from the life table referred to in option (1).

The latter choice is acceptable if the trend of the mortality rates for ages under 5 years agrees with the trend for older ages.

**Program BESTFT**

The procedure used in MORTPAK-BESTFT is slightly different. The United Nations program selects one of the United Nations model life tables that "best fits" the empirical data and, by using "least square criteria," determines a smoothed set of the $q_x$ values of the life table. The results of this program are three sets of the smoothed $q_x$ function, corresponding to the one, two, and
three component techniques developed by the United Nations for constructing model life tables. Each of these sets is different, and none of them reproduces the empirical infant mortality rate. If a life table is desired, the spreadsheet LTMXQXAD developed by the Bureau of the Census or the program LIFTB of MORTPAK can be used.

Constructing a Life Table from Fragmentary Data

In demographic analysis, situations sometimes arise in which life tables are required but no reliable information is available to construct them. By applying some of the techniques described in the text, fragmentary data of acceptable quality can be used to derive the information needed to construct a life table. For example, an estimate of life expectancy at any age can be derived using the Preston-Bennett or Bennett-Horiuchi technique; and estimates of infant and early childhood mortality can be derived by techniques using data on average number of children ever born and children surviving per woman, by age of mother. Based on these estimates, and assuming that the population to which the fragmentary information pertains has a particular pattern of mortality (from models or from another population with information), life tables can be constructed using two microcomputer programs developed by the United Nations. The programs are BESTFT and MATCH.

If more than one piece of data is available, the program to be used is BESTFT. If only one piece of data is available, either BESTFT or MATCH can be used. BESTFT requires one or several values of the \( q_x \), while MATCH can use \( q_x \), \( m_x \), \( l_x \), or life expectancy at any age, but only one value at a time.

Software

Documentation of the Bureau of the Census spreadsheets LOGITLX and LOGITQX are presented in volume II. See BESTFT and MATCH in the United Nations MORTPAK package.
REFERENCES


Chapter IV

FERTILITY

This Chapter in Brief

The purpose of this chapter is to present several procedures to measure the level of fertility in a population. As in the case of mortality in the previous chapter, the procedures for calculating the level of fertility depend on the availability of data and on the detail of the information. For cases where vital registration is complete, fertility can be measured directly using classical indices. Unfortunately, most developing countries do not have reliable vital statistics, and hence techniques have been developed to measure fertility indirectly based on census or survey information.

Direct estimation

The most frequently used indices are presented:

1. Crude birth rates and their standardization (for estimating the impact of changing age structure on fertility).
2. Age-specific fertility rates and total fertility rates and their standardization (mainly for analyzing changes in marital fertility and proportion of women married).

Indirect estimation

The chapter presents some of the techniques most frequently used to estimate fertility levels indirectly. These techniques are:

1. Techniques using the age structure of the population, such as:
   (a) Rejuvenation techniques, Bureau of the Census spreadsheet REVCBR.
   (b) The own-children technique.
   (c) Thompson technique.
   (d) Rele technique, Bureau of the Census spreadsheet RELEFERT.
2. Techniques based on special questions related to fertility, such as children ever born, first births, and births occurring during the 12-month period prior to a census or survey. These techniques are:
   (a) Brass P/P ratio technique and its modification by Trussell, Bureau of the Census spreadsheet PFRATIO.
(b) Brass $P_1/F_1$ ratio technique for first births.

(c) Arriaga technique, Bureau of the Census spreadsheets ARFE-2 and ARFE-3.

(d) Relational Gompertz technique.

**Spreadsheets and Methods That Can Be Used for Analyzing Fertility, According to the Available Information**

<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female population in reproductive ages by 5-year age groups, total population and:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Total fertility rate</td>
<td>CBR-CFR</td>
<td>Estimates the crude birth rate and the general fertility rate.</td>
</tr>
<tr>
<td>c. General fertility rate</td>
<td>CBR-TFR</td>
<td>Estimates the crude birth rate and the total fertility rate.</td>
</tr>
<tr>
<td>d. Pattern of fertility</td>
<td>ADJASFR</td>
<td>Adjusts the pattern of fertility proportionally to obtain a desired number of births.</td>
</tr>
<tr>
<td>Population age structure by 5-year age groups and mortality</td>
<td>REVCBR</td>
<td>Estimates the crude birth rate for years prior to census.</td>
</tr>
<tr>
<td></td>
<td>RELEFERT</td>
<td>Estimates the total fertility rate.</td>
</tr>
<tr>
<td>Average number of children ever born per woman and a pattern of fertility by age of women</td>
<td>PFRATIO</td>
<td>Estimates age-specific fertility rates.</td>
</tr>
<tr>
<td></td>
<td>ARFE-2</td>
<td>Estimates age-specific fertility rates.</td>
</tr>
<tr>
<td></td>
<td>ARFE-3</td>
<td>Estimates age-specific fertility rates.</td>
</tr>
</tbody>
</table>
Spreadsheets and Methods That Can Be Used for Analyzing Fertility, According to the Available Information--Continued

<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total fertility rates</td>
<td>REL-GMPZ</td>
<td>Estimates age-specific fertility rates.</td>
</tr>
<tr>
<td></td>
<td>ASFRPATT</td>
<td>Estimates age-specific fertility rates.</td>
</tr>
<tr>
<td></td>
<td>TFRLOGST</td>
<td>Interpolates and extrapolates total fertility rates (logistic function).</td>
</tr>
<tr>
<td></td>
<td>TFRSINE</td>
<td>Interpolates and extrapolates total fertility rates (sine function).</td>
</tr>
</tbody>
</table>

Where to Find the Software

The following spreadsheets for measuring the level of fertility in a population can be found in volume II.

Bureau of the Census spreadsheets:

<table>
<thead>
<tr>
<th>ADJASFR</th>
<th>REL-GMPZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARFE-2</td>
<td>RELFERT</td>
</tr>
<tr>
<td>ARFE-3</td>
<td>REVGBR</td>
</tr>
<tr>
<td>ASFRPATT</td>
<td>TFR-GFR</td>
</tr>
<tr>
<td>CBR-GFR</td>
<td>TFRLOGST</td>
</tr>
<tr>
<td>CBR-TFR</td>
<td>TFRSINE</td>
</tr>
<tr>
<td>PFRATIO</td>
<td></td>
</tr>
</tbody>
</table>
Introduction

Like mortality, fertility has begun to decline in many developing countries in recent decades, but in only some has the decline been as striking as the decline of mortality. Thus, in spite of significant reductions, birth rates (the number of births per 1,000 population) in developing countries remain substantially higher than those in developed countries (figure IV-1).

To formulate or evaluate policies concerning population growth, information is needed not only on the number of births, but also on trends in the number, proportion, and ages of women having births. This chapter describes some commonly used techniques to estimate levels and patterns of fertility. It begins with some conventional measures that can be calculated directly from registered data on births and population. Then, since many developing countries do not have complete or reliable vital registration systems, some techniques are presented to estimate fertility indirectly from deficient or incomplete data and data on particular characteristics. These include techniques based on the age structure of the population and on census data in which women report the number of children they have had.

Figure IV-1. Crude Birth Rates for Major Regions of the World

![Crude Birth Rates for Major Regions of the World](chart)

Direct Estimation of Fertility

Based on information on births and population, several indices can be calculated for measuring fertility and reproduction. Such information, which is not always free of errors, is provided by vital registers, censuses, and surveys. One of the most frequently used indices is the crude birth rate, which is directly related to natality and population growth. Other indices, such as the general fertility rate, age-specific fertility rates, and the total fertility rate, are used for measuring fertility levels and reproduction. Indices less frequently used or analyzed are the gross reproduction rate and the net reproduction rate; both of these are closely related to the concept of reproduction, or "replacement" of the population. Each of these indices is explained below.

Crude birth rate (CBR)

The crude birth rate, or the number of infants born in a year per 1,000 persons in a population, is calculated as the number of births occurring in a year divided by the population at midyear, times 1,000.

For example, the CBR for Hong Kong in 1987 is obtained as follows:

\[
\frac{1,000 \times 69,811}{5,613,000} = 12.44
\]

There were 12 births per 1,000 population in Hong Kong in 1987.

The crude birth rate is the most frequently used measure of fertility not only because it is easy to understand, but also because it requires the least amount of information. It indicates the growth of population without considering loss through mortality or migration.

When interpreting values of the CBR, it should be taken into account that the base for its calculation involves the whole population, including men, children, and women outside the reproductive ages. Hence, the level of the crude birth rate depends not only on the number of births, but also on the proportion of persons who are not subject to having children. Because it is affected by the sex and age structure of the population, this index is considered a "crude" measure. Two populations may have different crude birth rates even if the frequency of having children among women of reproductive ages is the same in each of them. This would occur when the age and sex compositions of the populations are different: the crude birth rate would be higher in the population in which women of reproductive age comprise a larger proportion of its people. Although the crude birth rate properly reveals the number of births per 1,000 population in a year, it fails to confine the measurement of births to women in reproductive ages. Several indices that do so are presented below.
General fertility rate (GFR)

The simplest measure that limits the number of birth to women of childbearing age is the general fertility rate, or the number of births in a year per 1,000 women ages 15 to 49 years.

For example, the GFR for Hong Kong in 1986 is obtained as follows:

\[
\frac{1,000 \times 72,221}{1,469,300} = 49.2
\]

(births) (women ages 15 to 49)

There were 49 births per 1,000 women of reproductive age in Hong Kong in 1986.

Although the general fertility rate represents a refinement over the crude birth rate, it still has its limitations. The frequency of births varies by age of women within the span of reproductive ages, and so populations in which women have the same frequency of birth at each age may have different general fertility rates due to differing age structure of women within the reproductive ages. Still other indices were developed to avoid the effect of different age compositions of the female population in reproductive ages. They are discussed below.

Age-specific fertility rates (ASFR)

An age-specific fertility rate is calculated as the number of births in a year to mothers of a specific age per woman (or per 1,000 women) of the same age at midyear. ASFR's are usually calculated for women in each 5-year age group for ages 15 to 49 years. (Although they may also be calculated for single years of age, this is rarely practical in developing countries.)

Age-specific fertility rates follow a fairly standard pattern among women in all populations: rates start from zero at very young ages, rising to a peak sometime in the twenties, then declining gradually until again reaching zero around 50 years of age. Slight variations to the pattern occur, depending on differences in age at marriage, on the proportion of women sexually active (mostly within marital unions), or on the desire and possibility of controlling pregnancies (mostly by using contraception). Relative declines in fertility rates are usually greatest for women under 25 years and over 35 years but, in absolute terms, largest declines tend to be among women in their late twenties and early thirties. Age patterns of fertility for selected countries in developing regions are illustrated in figure IV-2, and changes over time in a single country are presented in figure IV-3.
Age-specific fertility rates are calculated as the number of births to mothers in a particular age group in a year per 1,000 women (or per woman) in the same age group at midyear.

In symbols:

\[ n^t_{fx} = \frac{n^t_{Bx}}{n^t_{px}} \times 1,000 \]

Where:

- \( n^t_{fx} \) is the age-specific fertility rate for women between ages \( x \) and \( x+n \) for year \( t \);
- \( n^t_{Bx} \) is the number of births to women between ages \( x \) and \( x+n \) in year \( t \); and
- \( n^t_{px} \) is the number of women between ages \( x \) and \( x+n \) in year \( t \).

Figure IV-2. Age-Specific Fertility Rates for Selected Countries

Although ASFR’s properly measure the fertility of women in each age group, it is difficult to use them to make comparisons among populations or within a certain population over time. In addition, they do not easily portray the overall level of fertility. For these reasons, a summary index was developed, known as the total fertility rate.

**Total fertility rate (TFR)**

The total fertility rate is a summary measure independent of the age and sex composition of a population. It represents the average number of children a group of women would have by the end of their reproductive years if they had children according to a set of age-specific fertility rates pertaining to a particular year. In other words, if a group of women have been exposed to a given set of ASFR’s from age 15 to age 50, the average number of children they would have by age 50 is the total fertility rate.
The TFR is derived by cumulating the age-specific fertility rates (per woman) for all ages of women. When rates are calculated for the seven conventional 5-year age groups, the TFR is the sum of the rates for each age group, multiplied by five (the width of the age-group interval). Total fertility rates for world regions are presented in figure IV-4. An example of the calculation of age-specific fertility rates and the total fertility rate is presented in table IV-1.

The TFR can also be interpreted as the number of children that, on average, will replace each woman by age 50 if none of the children die. The concept of replacement is further explored in the gross and net reproduction rates.

Figure IV-4. Total Fertility Rates for Major Regions of the World

Table IV-1. Age-Specific Fertility Rates and Total Fertility Rate for Chile: 1983

<table>
<thead>
<tr>
<th>Age of women</th>
<th>Female population</th>
<th>Number of births</th>
<th>Fertility rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4) = (3)/(2) x 1,000</td>
</tr>
<tr>
<td>15-19</td>
<td>593,262</td>
<td>36,784</td>
<td>62.0</td>
</tr>
<tr>
<td>20-24</td>
<td>587,076</td>
<td>81,213</td>
<td>138.3</td>
</tr>
<tr>
<td>25-29</td>
<td>505,362</td>
<td>65,236</td>
<td>129.1</td>
</tr>
<tr>
<td>30-34</td>
<td>424,186</td>
<td>37,506</td>
<td>88.4</td>
</tr>
<tr>
<td>35-39</td>
<td>385,749</td>
<td>17,532</td>
<td>45.4</td>
</tr>
<tr>
<td>40-44</td>
<td>325,105</td>
<td>4,929</td>
<td>15.2</td>
</tr>
<tr>
<td>45-49</td>
<td>266,575</td>
<td>512</td>
<td>1.9</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td>480.4</td>
</tr>
<tr>
<td>Sum x 5 / 1,000 =</td>
<td></td>
<td></td>
<td>2.4</td>
</tr>
</tbody>
</table>

The total fertility rate in Chile in 1983 was 2.4 births per woman.

Gross reproduction rate (GRR)

The gross reproduction rate is a measure analogous to the total fertility rate, but it refers only to female births. Thus, it is derived in the same manner as the TFR but uses a set of age-specific fertility rates calculated based on female births only. As an acceptable approximation, it can also be derived by multiplying the TFR by the proportion of all births that are female. The GRR is usually interpreted as the average number of daughters that would replace each woman in the absence of female mortality from birth through the childbearing years, based on a given set of age-specific fertility rates (see calculation in table IV-2).

As a fertility measure, the gross reproduction rate has no particular advantage over the total fertility rate. As suggested above, its principal utility is as an index of generational replacement: it indicates the extent to which a group of women would be "replaced" by female children if the women had children according to a given set of age-specific fertility rates. It is a "gross" measure of replacement because it assumes that none of the girls die before they reach the age of their mothers in the reproductive years. Of course, in actual populations, some daughters die before reaching childbearing ages. Another index was developed to take into account mortality of the daughters. It is the net reproduction rate, discussed below.
Table IV-2. Calculation of Selected Fertility Indices

<table>
<thead>
<tr>
<th>Age of woman</th>
<th>Female population births (1)</th>
<th>Total Female births (2)</th>
<th>Female births (3)</th>
<th>ASFR for female births (4) = (2)/(1)</th>
<th>ASFR female births (5) = (3)/(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>439,921</td>
<td>29,607</td>
<td>14,533</td>
<td>.0673</td>
<td>.0330</td>
</tr>
<tr>
<td>20-24</td>
<td>396,227</td>
<td>64,297</td>
<td>31,351</td>
<td>.1623</td>
<td>.0791</td>
</tr>
<tr>
<td>25-29</td>
<td>313,238</td>
<td>52,031</td>
<td>25,441</td>
<td>.1661</td>
<td>.0812</td>
</tr>
<tr>
<td>30-34</td>
<td>249,856</td>
<td>34,089</td>
<td>16,827</td>
<td>.1364</td>
<td>.0673</td>
</tr>
<tr>
<td>35-39</td>
<td>205,264</td>
<td>19,799</td>
<td>9,692</td>
<td>.0965</td>
<td>.0472</td>
</tr>
<tr>
<td>40-44</td>
<td>170,106</td>
<td>8,181</td>
<td>4,021</td>
<td>.0481</td>
<td>.0236</td>
</tr>
<tr>
<td>45-49</td>
<td>137,970</td>
<td>1,385</td>
<td>700</td>
<td>.0100</td>
<td>.0051</td>
</tr>
<tr>
<td>Sum</td>
<td>1,912,582</td>
<td>209,389</td>
<td>102,565</td>
<td>.6867</td>
<td>.3367</td>
</tr>
</tbody>
</table>

General fertility rate

\[ \frac{209,389}{1,912,582} = .0043 \]

Total fertility rate

\[ 5 \times .6867 = 3.434 \]

Gross reproduction rate

\[ 5 \times .3367 = 1.683 \]

Approximation of the gross reproduction rate

\[ \frac{102,565/209,389} \times 3.434 = 1.682 \]

NOTE: ASFR refers to age-specific fertility rates.

Source: U.S. Bureau of the Census.
Net reproduction rate (NRR)

The net reproduction rate index was developed for theoretical demography rather than for practical population issues. Given the fact that women are the ones who give birth, it is interesting to know the average number of daughters that will replace each woman in childbearing ages in the next generation, in a population without migration. To determine the replacement index, mortality must be introduced into the calculation. The methodology assumes that both fertility and mortality remain unchanged in the future. Data needed to calculate the NRR are fertility rates and a life table providing survivorship values from birth to the age of the woman in the childbearing years. The steps in the calculation are as follows:

1. Multiply each age-specific fertility rate pertaining only to female births by the probability of surviving from birth to the age to which the fertility rate pertains. The survivorship ratio is the $nL_x$ value from an appropriate life table, divided by 100,000 times the age interval (n). The age intervals of the $nL_x$ values and of the fertility rates are usually 5 years, hence they are divided by 500,000; and the age group of the fertility rate, x to x+n, represents the age of the woman.

2. Sum the products obtained in step (1).

3. Multiply the sum by 5.

The calculation can be reduced to steps (1) and (2) when the age intervals are 5. In this case, divide the $nL_x$ function by the radix of the life table (100,000), but do not multiply by the age interval (see table IV-3).

A net reproduction rate equal to 1.0 is often referred to as "replacement level fertility" because it indicates that each woman will be replaced by exactly one other woman after a generation. Under these conditions, if fertility and mortality remain constant at the levels that produced the NRR of 1.0, the population will eventually stop growing if there is no migration. A value of the NRR sustained above 1.0 indicates that, on average, more than one woman will replace each woman in the childbearing ages after a generation, and hence the population will continue to grow. A value of the NRR sustained under 1.0 (currently the case in some European countries) indicates that, on average, "less than one" woman will replace each woman in the childbearing ages after a generation, and hence the population will eventually decline.
Table IV-3. Calculation of the Net Reproduction Rate

<table>
<thead>
<tr>
<th>Age of woman</th>
<th>Female population</th>
<th>Female births</th>
<th>ASFR for female births</th>
<th>100000 nlx times</th>
<th>col.(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>439,921</td>
<td>14,533</td>
<td>.0330</td>
<td>4.36504</td>
<td>.1442</td>
</tr>
<tr>
<td>20-24</td>
<td>396,227</td>
<td>31,351</td>
<td>.0791</td>
<td>4.28169</td>
<td>.3388</td>
</tr>
<tr>
<td>25-29</td>
<td>313,238</td>
<td>25,441</td>
<td>.0812</td>
<td>4.16710</td>
<td>.3384</td>
</tr>
<tr>
<td>30-34</td>
<td>249,856</td>
<td>16,827</td>
<td>.0673</td>
<td>4.02657</td>
<td>.2712</td>
</tr>
<tr>
<td>35-39</td>
<td>205,264</td>
<td>9,692</td>
<td>.0472</td>
<td>3.85234</td>
<td>.1819</td>
</tr>
<tr>
<td>40-44</td>
<td>170,106</td>
<td>4,021</td>
<td>.0236</td>
<td>3.62483</td>
<td>.0857</td>
</tr>
<tr>
<td>45-49</td>
<td>137,970</td>
<td>700</td>
<td>.0051</td>
<td>3.29859</td>
<td>.0167</td>
</tr>
</tbody>
</table>

Net reproduction rate is the sum of col(5) 1.3769

Note: The survival ratios are from the life table presented in table III-3.

Source: U.S. Bureau of the Census.

Consistency Among Fertility Rates

The various fertility rates for a given population are interrelated. The total fertility rate and the gross reproduction rate differ only in that the latter is confined to female births. Similarly, once a female age structure in reproductive ages, a set of age-specific fertility rates, and the total population (both sexes) are given, then the crude birth rate, the total fertility rate and the gross reproduction rate must be compatible. It is advisable to verify that independent estimates of fertility rates are consistent based on an age structure and a pattern of fertility rates.

In light of the need for such consistencies, the Bureau of the Census has developed a set of spreadsheets that can be used for estimating or calculating some fertility rates, provided that other rates are available. These spreadsheets are: ADJASFR, TFR-GFR, CBR-GFR, AND CBR-TFR. They are described in appendix IV-1, and their documentation is presented in volume II.
Standardization of Crude Birth Rates

The crude birth rate, it will be recalled, is affected by the age and sex composition of the population. Thus, it is difficult to compare birth rates among two or more populations with different age structures or within a particular population over time. While indices measuring changes of fertility, such as the TFR, GRR, and NRR, are not affected by the age structure, neither do they reflect the actual impact that age structure has on population growth. Since birth rates are more directly related to population growth than fertility indices are, it is useful to analyze changes in the crude birth rate to determine if they are due to changing age structure or to changing fertility. This analysis can be done by using a standardization process. As discussed earlier in the case of the crude death rate (see chapter III, appendix III-1), the discrepancy between the crude birth rates of two populations is a combined product of differences between their age and sex compositions and differences between their fertility rates. As with the crude death rate, there are two methods of standardizing the crude birth rate: direct and indirect (Spiegelman, 1968).

Direct and indirect standardization of the crude birth rate eliminates the effects of the age structure for comparative purposes. Standardized crude birth rates differ mainly because of actual differences in fertility levels. In comparing two or more populations, indirect standardization reveals where fertility is higher or lower, even when fertility rates are not available for each of them. On the other hand, direct standardization requires that fertility rates be available for each population. If fertility rates are available, standardization is not needed to determine differences in fertility, but direct standardization is still useful to detect the impact of age structure and fertility levels on crude birth rates. Since standardization procedures for the CBR are similar to those for the CDR, they will be presented only in summary form. (See chapter III, Standardization of Crude Death Rates, and appendix III-1).

Direct standardization

The direct method requires the availability of information on age-specific fertility rates for all populations to be compared. After selecting a population to be used as the "standard," carry out the following steps:

(1) Multiply the age-specific fertility rates of each population by the number of females in the corresponding age group of the standard population.

(2) Sum the products derived in step (1) to obtain the total "expected" births based on the fertility in each population being studied and the population age structure taken as standard.

(3) Divide each sum by the total of the standard population to obtain the standardized birth rate pertaining to the fertility of each population being studied.
Each standardized birth rate is the crude birth rate that would be observed in the standard population if it had the age-specific fertility rates of the particular population under consideration. If several populations are being compared using the same standard, the higher the standardized rate, the higher the fertility of the population. Of course, as mentioned in the case of mortality, standardized rates are affected by the choice of standard population, creating difficulty in the interpretation of their differences and limiting their utility.

**Indirect standardization**

The indirect method is used when information on age-specific fertility rates is not available for the populations being studied. Only a distribution of the population by age and sex and a total number of births are required for each population to be standardized. A population with information on the crude birth rate and age-specific fertility rates must also be available to serve as the standard. The procedure consists of the following steps:

1. Apply the standard age-specific fertility rates to the corresponding number of women in each population to be standardized, and take the sum of the resulting births by age of women. The result is the "standard" number of births for each population.

2. Divide each population's actual number of births by the standard number of births. The result in each case is an "adjustment factor" representing the difference between the fertility of the population to be standardized and that of the standard population.

3. Multiply the standard population's crude birth rate by each adjustment factor. This will increase or decrease the standard crude birth rate according to whether the fertility of each population is higher or lower than that of the standard.

The standardized rates computed by the indirect method provide a comparison of fertility between each population being studied and the standard population. However, when it is calculated for various populations using the same standard, it does not provide a direct comparison of fertility among the populations because each standardized rate depends on the age composition of the population in question. At best, fertility of the various populations may be ranked according to their relative standing in comparison to fertility of the standard population.

As in the case of mortality, the difference between two crude birth rates can be decomposed into the portions due to fertility, age structure, and interaction. (See chapter III, Decomposition of the difference between two crude death rates).
Standardization of Total Fertility Rates

Generally, the start of a decline in fertility from high levels coincides with other changes in the characteristics of a population. A change in social characteristics may have an impact on the age at which women enter marital unions. The changing attitudes toward entering marital unions will have an impact on the proportion of women who are mothers. A change in the proportion of mothers in reproductive ages will affect the number of children that the female population is having, and hence the fertility and the total fertility rate.

Total fertility rates are the sum of the age-specific fertility rates times 5, if such rates pertain to 5-year age groups. Since age-specific fertility rates can be decomposed into the proportion of females who are mothers and the fertility of those females who are mothers, the change in a total fertility rate can be decomposed into these same two components. This procedure provides a measure of how much of the fertility change is due to a change in fertility among women having children, and how much to a change in the proportion of women who have children.

The required information is a set of age-specific fertility rates and the proportion of females who are mothers in the population, both by 5-year age groups. Sometimes, if the proportion of mothers is not available, the proportion of married women is taken as a proxy. In these cases, it should be taken into account that not all married women have children, and that not all women having children are married.

In an example presented in appendix IV-2, table A-IV-2.1, total fertility rates changed from 5.67 to 5.12. The change can be decomposed into what is due to the change of the proportion of women exposed to motherhood (or those who are married) and the part due to changing fertility of mothers. While fertility of mothers increased in the youngest age group, it tended to decline in other ages. The proportion of mothers declined considerably in the young adult ages and remained practically the same in the older reproductive ages. Overall, the changing fertility of mothers accounted for 51.5 percent of the change, and the changing proportion of mothers accounted for 48.6 percent. The remaining -0.2 percent was due to the change of both, or interaction.

Indirect Estimation of Fertility

Techniques presented in the last section were based mostly on data gathered in vital registration systems. Many developing countries do not have vital registration data (or census or survey data) that are complete enough or of sufficient quality to derive these direct measures; indeed, some countries do not have any vital registration systems at all. For this reason, researchers have developed various techniques to estimate fertility indirectly. Some of the methods use data gathered for other or general purposes, such as information on age and sex collected in a census. Others use data tabulated from special questions included on census or survey questionnaires, such as information on how many children each woman has had,
commonly known as data on "children ever born" per woman. Both types of techniques are discussed in the following sections.

Techniques based on enumerated population

As discussed in chapter II, information on the age structure of a population is like a picture of the past demographic events that occurred in the population. The population enumerated in a certain age group in a census represents the survivors of a number of infants born as many years ago as the age of the group, after exposure to mortality and migration during the interim. Hence, the population under age 5 enumerated in a census represents the survivors of infants born during the 5 years prior to the census, if there was no migration (migration will affect this age group only slightly in most cases). If the mortality of the survivors from their birth to the census date is known, then the total number of births during the 5 years prior to the census can be estimated. Having the total number of births for a 5-year period, the birth rate can be estimated. This technique for estimating the birth rate is known as "rejuvenation."

Some other techniques are also based on the age structure of the population. One of them requires additional information on the mothers of the children, known as the "own-children" technique. Others, developed by Thompson (U.S. Bureau of the Census, 1931) and Rele (1967), use the ratio of children to women for estimating the total fertility rate. These techniques are discussed below.

Rejuvenation technique. The chapter on mortality introduced the concept of survival ratios (the $n_p$ function of the life table), representing the proportion of persons in a given age group who will survive to an older age group, during a specified period of time. Survival ratios are often used to project a population, for example, to calculate the number of persons age 20 to 24 years who will live to be 25 to 29 years. This is done by multiplying the population age 20 to 24 by the corresponding survival ratio, with the result representing the population age 25 to 29 exactly 5 years later if there has been no in or out migration of persons of these ages.

Survival ratios may also be used to "reverse survive" the population in selected age groups: in this case, the population age 25 to 29 at the later date is divided by the survival ratio, with the result representing the persons age 20 to 24 years at a date 5 years earlier, from which the 25 to 29 year olds have survived. The process of reverse survival is commonly known as "rejuvenation" and can be used as a technique to estimate fertility. This is done by rejuvenating the population at young ages to calculate the number of births of which they are the survivors.

To apply the rejuvenation technique, it is necessary to have information on the population by age (5-year age groups are sufficient), reliable life tables representing the mortality of the population during the past 5, 10, or 15 years, and estimated migrants by age, if migration is significant. While the population and survival ratios are frequently available by sex, the
The rejuvenation technique may also be applied for both sexes combined (see appendix IV-3).

The technique assumes that completeness of census coverage was the same for children as for the rest of the population.

Although the rejuvenation technique can theoretically provide a crude birth rate for a single calendar year, it is usually used to estimate the crude birth rate for two or three 5-year periods preceding the census date. When used in this way, different life tables should be used for the two periods if mortality is changing. The method is applied as follows:

(1) Rejuvenate the population at ages under 15 years for three 5-year periods prior to the census date. Do this in two successive steps by dividing the population in each age group (for either sex or both sexes) by its corresponding survival ratio. The results for the population ages 0 to 4 years, 5 to 9 years, and 10 to 14 years at the census date represent births that occurred during each of the three 5-year periods preceding the census date.

(2) Next, an estimate of the midperiod population is needed for the three 5-year periods. These populations can be estimated by using the total enumerated population and growth rates for years prior to the census.

(3) Now the crude birth rates can be calculated. First, calculate the average annual number of births for each of the three 5-year periods by dividing the total births calculated in step (1) by five. Then, for each period, divide the average annual number of births by its corresponding midperiod population as estimated in step (2), and multiply by 1,000. The results represent the number of births per 1,000 population for the three 5-year periods prior to the census date as estimated by the rejuvenation technique.

This method is attractive because it does not require the collection of any data related specifically to fertility. The reliability of the estimate depends primarily on the quality of the census data by age and on the quality of the life tables used to represent mortality during the three 5-year periods prior to the census date. In many censuses, enumeration is less complete for ages 0 to 4 years than for other ages, with the result that the estimate based on these ages may underestimate the crude birth rate. It is sometimes preferable to accept the estimate based on ages 5 to 9 years or 10 to 14 years, but these values are more dependent on the mortality data used and refer to dates 7.5 and 12.5 years prior to the census. Thus, if fertility is changing, they may not reflect the level of the crude birth rate at the time of the census. See description of the Bureau of the Census spreadsheet REVCBR in appendix IV-3 and its documentation in volume II.

Own-children technique. Under certain circumstances, census data by age can be used to obtain not only a crude birth rate, but age-specific fertility rates as well, for periods up to 10 or 15 years preceding the census date (Grabill, 1942a; Grabill and Cho, 1965; Cho, 1973; and Retherford and Cho, 1973).
The own-children technique uses census information on children and women by single years of age. In addition, information individualizing the children in relation to their natural mothers facilitates the calculations and avoids having to make certain assumptions. The idea behind the technique is a rejuvenation process. If the current age of the mother and of her children is known, it is possible to calculate not only the year of birth of each child, but also the age of the mother at the birth of the child. With this information, an estimate of the number of births for each year prior to the census can be obtained, as well as the distribution of births by age of mother. Since a census also provides data on all females by single ages, they can be rejuvenated for several years prior to the census. With data on children by age of their mothers and females by single ages for several years prior to the census, age-specific fertility rates can be calculated. The method takes into account mortality of both women and children during the rejuvenation period.

Although census questionnaires nearly always include the required questions, special tabulations must usually be made to obtain the requisite data. To apply the own-children technique, the following data are needed:

1. The number of children under 15 years of age living with their own mothers ("matched children"), classified by both their own single years of age and their mothers' single years of age.

2. The number of children under 15 years of age not living with their own mothers ("unmatched children"), classified by their own single years of age.

3. The total number of women, by single years of age.

4. Appropriate life table survivorship ratios for up to 15 years prior to the census date.

Implicit in the methodology of the own-children technique are the following assumptions:

1. Census coverage is the same for children as for women.

2. The age distribution of mothers of unmatched children is the same as the age distribution of mothers of matched children.

When the technique is applied, unmatched children are first distributed by age of mother according to the pattern available for matched children. To calculate age-specific fertility rates for recent periods in the past, the number of births and their distribution by age of mother, and the number of women by age are needed. To obtain these data, the technique rejuvenates census data on (a) children by age of mother and (b) all women. During the rejuvenation process, each child's own age and the age of its mother remain linked to the child. The computer program used to apply the own-children technique rejuvenates the children by age of mother in such a way that the result indicates in which year the children were born and the age of the mother at that time; in other words, it presents for each year a distribution
of births by age of mother. A separate rejuvenation process, based on the census distribution of all women by single years of age, provides the number of women (not just mothers) at each reproductive age for each year of the rejuvenation period.

Once the rejuvenation has been carried out, the number of births by age of mother for each year is divided by the corresponding female population by age, multiplied by 1,000. The result is a set of age-specific fertility rates for single years of age based on the own-children technique.

In many cases, these single-year rates form an implausible pattern, due primarily to errors in age reporting. Two adjustments can be made to smooth the pattern: the rates may be grouped from single-year rates into rates corresponding to 5-year age groups of women, and the data may be combined into rates for 3-year or 5-year periods of time.

A major advantage of the own-children technique is that it does not require any assumption about the trend of fertility (for example, that fertility remained constant in the past). In fact, the technique can be used to estimate fertility trends and patterns.

The own-children technique is not without its problems, however. In any country, there will be children whose mothers do not live with them in the same household, and there will be mothers in the household whose children live elsewhere. Children in the first instance represent the unmatched children of the household, while those in the second instance are likely to be found as unmatched children in some other household. Although it is assumed that the ages of mothers of unmatched children are the same as those of matched children, this often may not be true.

Accuracy of the results is further affected by age misreporting and differential census coverage by age. For example, if there is age heaping at ages 5 and 10 years, fertility rates may be overestimated for exactly 5 and 10 years prior to the census. If children are undercounted to a greater extent than adults, fertility rates may be underestimated, especially for the years immediately prior to the census. See appendix IV-4 for further discussion of the own-children technique.

**Thompson technique.** Based on the fact that most of the children enumerated in a census are related to the enumerated women in childbearing ages, Thompson (U.S. Bureau of the Census, 1931) derived a simple technique for estimating the net reproduction rate in a population. The ratio of female children under age 5 to women in childbearing ages is, in a certain way, a fertility index, since it approximates the average number of female children born per woman in the childbearing ages during the 5 years prior to the census. Using stable population theory, Thompson reasoned that if such a ratio is divided by a similar ratio from a female life table which has a net reproduction rate of 1.0, the quotient represents the net reproduction rate of the actual population (see table IV-4).
Table IV-4. Calculation of the Thompson Index for Estimating the Net Reproduction Rate

<table>
<thead>
<tr>
<th>Age</th>
<th>Female population</th>
<th>nLx function</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>654,258</td>
<td>451,973</td>
</tr>
<tr>
<td>5-9</td>
<td>503,070</td>
<td>445,800</td>
</tr>
<tr>
<td>15-49</td>
<td>1,560,666</td>
<td>2,761,616</td>
</tr>
<tr>
<td>20-54</td>
<td>1,376,847</td>
<td>2,612,137</td>
</tr>
</tbody>
</table>

Index based on ages 0-4 and 15-49

\[
\text{NRR} = \frac{654,258}{1,560,666} = 2.561474
\]

Note: The population is from table II-1, and the nLx values from the life table of table III-3.

Source: U.S. Bureau of the Census.
Rele technique. The Rele (1967) technique also is based on the observation that the ratio of children to women is related to a certain fertility level. For instance, in a population, the ratio of infants under age 1 to all women in childbearing ages would be close to the general fertility rate. Infants under age 1 are the survivors of those born during the past year, and the female population in childbearing ages approximates the population having had the births during the past year (except, on the average, for a theoretical one-half year of age). If, instead of the population under age 1, the population under age 5 is taken, the relationship with fertility during the last 5 years will still be strong, although it will require some adjustment because of mortality during the period. Therefore, if the mortality level of a population is known, as well as a particular pattern of fertility, the relationship between certain ratios of children to women and the level of fertility has to be very close. Such a strong relationship would permit the estimation of any of the fertility indices that were presented earlier in this chapter.

The Rele technique estimates a total fertility rate for one or two 5-year periods prior to the census date based on data on children and women by age, and life expectancy at birth for both sexes. The data on age are used to calculate various child-woman ratios. The important aspect of these ratios is the age of the children. Ratios based on children ages 0 to 4 years provide estimates of fertility for the 5-year period prior to the information date, while ratios based on children ages 5 to 9 provide estimates of fertility for a period 5 to 10 years prior to the information date. The estimates are based on the following four ratios:

(1) The ratio of children age 0 to 4 to women age 15 to 44.
(2) The ratio of children age 0 to 4 to women age 15 to 49.
(3) The ratio of children age 5 to 9 to women age 20 to 49.
(4) The ratio of children age 5 to 9 to women age 20 to 54.

The computer program RELEFERT applies the Rele technique for converting the child-woman ratios to gross reproduction rates based on a set of coefficients that Rele developed. The coefficients are based on his observation that for any given level of mortality, there is a linear relationship between the child-woman ratio and the gross reproduction rate. Although the coefficients were established based on stable population theory, the relationships can be used in actual populations.

Once the gross reproduction rates have been derived, they are converted to total fertility rates by calculating the ratio of the number of total births to female births, and multiplying this ratio by the GRR.

An advantage of the Rele technique is that its use requires only an age distribution of the population and a life expectancy at birth. If all the child-woman ratios are used, the technique provides two estimated total
fertility rates for each of the two 5-year periods prior to the census date. All of the results should be examined for consistency among themselves as well as in relation to estimates derived by other methods before accepting any one as a final estimate of fertility. If the trend seems implausible or the rates are inconsistent, the age distributions used to calculate the child-woman ratios should be examined for any peculiarities. It should be kept in mind that the method is sensitive to differential underenumeration of children and women. If children are undercounted to a higher degree than women, which is often the case, the result will be an underestimation of the TFR. Likewise, if children age 0 to 4 years are undercounted to a higher degree than those age 5 to 9 years, the estimates may imply a spurious decline in fertility or a larger decline than is actually true. Research has shown that the method is less sensitive to errors in the data on life expectancy at birth. See appendix IV-5 and the documentation of spreadsheet RELEFERT in volume II.

Techniques based on special fertility questions

So far this chapter has discussed techniques to estimate fertility directly from vital registration data and indirectly from the age structure of the population. This section presents some techniques that can be applied in another situation of data availability: indirect estimation based on special questions on fertility included in censuses and surveys. Many census and survey questionnaires contain a fertility component asking women about the number of children they have ever had and whether they had a birth in the year preceding the inquiry. Several techniques have been developed to derive estimates of age-specific fertility rates based on these data.

Fertility rates are related to the average number of children ever born per woman. For example, if fertility were to remain constant, the average number of children ever born per woman in the reproductive ages would be similar to the cumulative fertility rates from the beginning to the end of the reproductive ages. In other words, the average number of children ever born per woman (or mean parity) at a given age represents the average cumulative fertility up to that age.

Based on this observation, Mortara (1949) used census data on children ever born per woman to estimate age-specific fertility rates. Working under the assumption that fertility was constant during the past, he graphically converted the data on children ever born from 5-year age groups to single ages of women by drawing a smooth line passing through the midpoint of each age interval of the data on children ever born. Such a graph represented not only the average number of children ever born by single ages of women, but also the cumulative fertility up to each age (because of the assumption of constant fertility during the past). Hence, he derived age-specific fertility rates for single ages by taking the differences between the average number of children ever born at successive ages. Mortara applied this process in a country where fertility was almost constant during the past, and where census information was of rather good quality. However, if fertility is changing and data are of poor quality, the estimates could be biased. In some populations, it is observed that women over age 40 may underreport the average number of children ever born, for various reasons. For example, there may be a reluctance to report dead children or those who already are living in other
households. Age of mother may also be misreported, giving a distorted pattern of fertility based on information on children ever born by age of mother.

Several years after the introduction of the Mortara procedure, other techniques were developed. Most of them use information on the average number of children ever born up to ages 30 or 35 to estimate the level of fertility. Hence, if an age pattern of the fertility rates is available, it can be adjusted to the level of fertility. The age pattern of fertility can be derived from any acceptable source of information, such as registers, surveys, or censuses. Since a large number of countries do not have a good vital registration system, questions concerning information on the pattern of fertility have often been included in the population census. A question about the number of births during a recent period of time (frequently 12 months) serves that purpose.

Three techniques for estimating fertility from such data are discussed below: the P/F (Parity/Fertility) ratio developed by Brass (Brass, et al., 1968), which uses the average number of children ever born to women in 5-year age groups and the pattern of fertility; the P1/F1 ratio technique, also developed by Brass (1975), which uses a similar rationale but determines the accepted fertility level based on first births only; and the Arriaga (U.S. Bureau of the Census, 1983) technique, which uses the same information as the P/F ratio technique but can be applied also in cases where the pattern of fertility is not available. Some adjustments to the basic data are required before applying each of these techniques; these are discussed below, and then the techniques are described.

Data problems and adjustments. To derive a pattern of fertility, information on births in the last 12 months by age of mother is used in most cases. Registration data may be also used if they are available or, alternatively, data collected in a census or survey that included a question on this topic may be used. The techniques assume that if there is a problem of under- or overreporting of children in the information used for estimating the pattern of fertility, such errors are relatively the same for all ages of mothers. In other words, they assume that the information represents the proper proportional distribution of the age-specific fertility rates.

If the information pertains to surveys or censuses, special attention must be paid to identifying the reference date of a woman's age: whether the reported age refers to her age at the date of the interview or her age at the time of the reported birth. Usually it refers to the date of the interview, and an adjustment is required to make it refer to the date of the reported birth so the age pattern of fertility will not be biased toward older ages. The adjustment is equal to approximately one-half of the reference period; if the data refer to births during the 12 months preceding the interview, each woman was, on average, one-half year younger at the time of the birth than at the time of the interview. This half-year adjustment is needed for all three of the techniques described below.
**P/F ratio technique.** This technique, developed by William Brass, adjusts an age-specific fertility pattern to a level of fertility derived from the information on children ever born. First, the pattern of fertility is cumulated up to ages 20, 25, ..., 50. As explained below, these cumulated fertility rates are adjusted (represented by the letter F) and they are compared with the children ever born, represented by the letter P (parity). Since in this technique the number of children ever born is accepted as representing the actual level of cumulative fertility, the P/F ratios for each age group are used to adjust the fertility pattern to the level indicated by the children ever born.

A ratio is calculated for each age group in the childbearing ages. Although all of these ratios can be used for adjusting the fertility pattern, most of the time only two of them are used: those for ages 20 to 24 and ages 25 to 29. The information required is the number of children ever born by age of mother (5-year age groups in the childbearing ages), the number of females in childbearing ages by 5-year age groups, and the number of infants born during a specific period of time prior to the census by age of their mothers (5-year age groups).

Before taking the P/F ratios, the information requires an adjustment. The average number of children ever born per woman refers to 5-year age groups whose central ages are 17.5, 22.5, 27.5, and so forth. On the other hand, the cumulative fertility refers to ages 20, 25, 30, and so forth. Since the ages do not match, an adjustment is required. Brass developed some adjustment factors for this purpose, simulating the pattern of fertility and the corresponding children ever born by using a third degree polynomial. Later, Trussell developed other adjustment factors using a set of fertility models (Coale and Trussell, 1974). Results from each of these adjustment factors are practically the same. The Bureau of the Census spreadsheet PFRATIO is available for calculating the Trussell variant (United Nations, 1983) of the P/F ratio technique (see appendix IV-6, and documentation in volume II).

The P/F ratio technique assumes that the completeness of data from which the age-specific fertility rates are calculated is the same for all age groups of women; that the reporting of the average number of children ever born per woman is complete at least up to ages 30 or 35 years; that there is no age misreporting of women in the childbearing years; and that the pattern and level of fertility have not changed during the 10 to 15 years prior to the census or survey.

Before deciding which of the P/F ratios to select, they should be analyzed. The P/F ratios themselves serve as a useful tool to analyze the quality of data and determine a possible recent trend in fertility. For instance, typical results may show similar P/F ratios for the age groups 20 to 24, 25 to 29, and 30 to 34 years, with ratios becoming smaller for the older ages. This could indicate that fertility has been constant in the past as assumed in the technique; that any underreporting of births in the last year occurred equally at all ages of women; and that any underreporting of children ever born occurred only with respect to women age 35 years and over.
Deviations from the typical results may indicate either violations of the assumptions or different patterns of underreporting. For example, a rising trend in the P/F ratios by age of women may suggest that fertility has been decreasing in the recent past. On the other hand, a declining trend in the P/F ratios by age of women may suggest either that fertility has been increasing or that reported data on children ever born suffer from progressively increasing omissions of children as age of women increases. Large fluctuations in P/F ratios may reflect either differential coverage by age or selective age misreporting of women, and again the results must be viewed with caution.

**Brass P1/Fl ratio technique.** The rationale behind this technique is similar to that of the P/F ratio technique, but this one is based on the premise that reporting of first births is probably more complete than reporting of other births (Brass, 1975). Thus, "P1" refers to the proportion of women who are mothers (that is, who have had at least one birth) and "Fl" refers to the proportion of women who have had a "first birth" in the last year. The values for P1 are derived from census or survey data on children ever born, while the values for Fl are calculated as age-specific "first birth" rates based on census or survey information on births in the 12 months preceding the interview. (A respondent who reports a birth in the last 12 months would also have reported one or more children ever born, and so it can be determined whether or not her recent birth was a first birth.) As noted above, it must be considered whether ages of women are recorded as of the date of the birth or as of the date of the interview, with appropriate adjustment if necessary. The "first birth" rates are cumulated up to successive age groups of women, and the result represents the proportion of women up to each age who would become mothers according to the first birth rates.

P1/Fl is a ratio of the proportion of women at a given age who are mothers (P1) to the cumulated "first birth" rate up to the same age (Fl). It serves as an estimate of the extent of underreporting (or overreporting) of first births. The technique assumes that the P1/Fl ratio is also appropriate for all births, and it is used as a factor to adjust the age-specific fertility rates according to an accepted pattern pertaining to births of all orders.

This method has several perceived advantages: it is thought that first births are more accurately reported than subsequent births, partly because they usually occur to younger women who tend to be better educated than older women and may therefore better understand the intended reference period in reporting births in the last 12 months. The technique places less demand on the data on children ever born as well, since it requires knowledge only about whether or not a woman has had any children, not the exact number of children she has had. Finally, the assumption regarding the pattern of fertility is less restrictive, requiring only that the age pattern of women having first births has not changed in the recent past. Thus, it is less affected than the overall P/F ratio technique by changes in marital fertility, especially in cases where contraceptive methods may be used to space second and subsequent births or to avoid having high-order births altogether.
The technique does have some disadvantages, however. The information on first births as calculated for this technique is not entirely independent of the information on children ever born, since the fact that a birth in the past 12 months was a first birth is determined by comparison with the response about children ever born. Furthermore, results of the technique may not be valid in situations where age at marriage, and consequently age at first birth, are changing rapidly. Finally, it may not always be appropriate to assume that the adjustment factor based on first births applies to all births. For all of these reasons, the results should be interpreted with caution.

Arriaga technique. One of the limitations of the previous techniques is that they are based on the assumption that fertility has been constant during a certain period in the past. Arriaga (U.S. Bureau of the Census, 1983) developed a technique for which such an assumption is not required. Based on a simulation model, he observed that under conditions of declining fertility, the number of children ever born by age of mother changes almost linearly for mothers' ages under 35 years. Based on this observation and on the fact that the reported number of children ever born by mothers under age 35 is usually acceptable, linear interpolation of the data on children ever born per woman by age of mother from two or more censuses can provide an estimate of the children ever born for 1 year prior (or posterior) to the date of the census. Hence, having information on the average number of children ever born per woman by age of mother for 2 consecutive years, the cohort differences between them for each single year of age of the female population represent the age-specific fertility rates by single years of age. These single-year age-specific fertility rates are affected by a possible decline in fertility, and an analysis of them detects the change. Thus, the technique does not require an assumption of constant fertility.

Misreporting of children in older ages of the female population would produce unacceptable fluctuations in the fertility pattern. However, as in the case of the previous techniques, if an age pattern of fertility is available, such a pattern can be adjusted to the fertility level implied by the fertility rates derived from the information on children ever born. The advantage in this case is that no adjustment is required when comparing children ever born with the cumulative pattern of fertility because in this technique the comparison is made between two sets of cumulative fertility rates. One set represents the pattern and the other is derived from the data on children ever born and represents the level of fertility.

The Arriaga technique also can be used when information on the average number of children ever born by age of mother and pattern of fertility are available for only one date. In this case, it assumes that fertility was constant during the past, and the results are practically the same as in the previous two techniques. If the pattern of fertility is not available, the technique estimates fertility based only on children ever born, but the results are more sensitive to errors in the data.

The principal advantage of the Arriaga technique over the standard P/F ratio technique is that it does not require an assumption of constant
fertility, and thus, when it is applied in populations where fertility is declining, results may be more reliable.

This technique, like the two previous ones, assumes that completeness in the recording of births in the last 12 months is the same for all age groups of women; that the average number of children ever born per woman is reported accurately for women under 30 or 35 years of age; and that there has been no age misreporting of women in childbearing ages. In addition, this technique assumes that changes in fertility will produce a linear change in the average number of children ever born per woman at each single year of age of women between the two reporting dates.

Fertility rates derived from data on the average number of children ever born by age of mother are sensitive to possible age misreporting of mothers, which may produce some problems in analyzing the results. In a certain way, however, this may be an advantage because it encourages an evaluation of the quality of the information being used. The factors used for adjusting the pattern of fertility to the level of fertility (the $Z_i$ factors in the Arriaga technique) are related not to changes in fertility but to the quality of data. If these factors are close to one, they indicate a consistency between the fertility pattern and the children ever born. If the factors differ from one but they are of similar magnitude, they indicate that the information is comparable and that the pattern of fertility has to be adjusted according to the proper level. But if the $Z_i$ factors show large fluctuations by age, they indicate that the information is not comparable and that there are inconsistencies in the information used. In this case, it is recommended to use the technique for each date separately, to analyze the results, and to be cautious about any conclusions because the data are not comparable.

The adjustment factors $Z_i$ are also related to the proper age of the mother at the moment of birth. With accurate data, these factors should be close to one for the pattern of fertility reflecting the age of the mother at birth. For instance, if the census data are accurate and the age of the mother was declared as of the census date while births refer to a 12-month period prior to the census, the $Z_i$ factors would fluctuate less with the pattern of fertility by age of mother as adjusted by a half year. But if the age of the mother was reported as of the moment of the birth, then the $Z_i$ factors will be closer for the unadjusted pattern of fertility. However, due to errors in age misreporting or misunderstanding of the period of time used to obtain the pattern of fertility, it is advisable to analyze such $Z_i$ factors from both the adjusted and the unadjusted pattern of fertility. In these cases, the user may choose the estimates where the adjustment factors $Z_i$ are more similar. Nonetheless, the user should be aware that the information has errors, and the results should be used with caution.

There are two spreadsheets that estimate fertility based on this technique. One of them, ARFE-2, should be used in cases where information on children ever born and the pattern of fertility is available for one or two dates. The other program is called ARFE-3, and should be used if information is available for three dates. See appendix IV-7, and documentation in volume II.
Relational Gompertz Technique. This technique estimates total fertility rates based on the information on (a) the average number of children ever born by age of mother, and (b) the pattern of fertility (Brass, 1981). Once the total fertility rate has been estimated, any available pattern can be adjusted to the estimated level as measured by the total fertility rate.

The technique, developed by Brass (1981) uses the Gompertz function to estimate fertility. The Gompertz function follows closely the pattern of the cumulative fertility rates. Hence, because of the similarity of the cumulative fertility rates to the average number of children ever born, the Gompertz function can be fitted to both.

The estimation process can be based on either the average number of children ever born by age of mother or the age-specific fertility rate for each age group. The author developed two standards for comparison with the actual data to obtain the estimates. The use of the standards was needed because the Gompertz function is fitted to the proportions of the total fertility rate cumulated up to each age group. Use of the standard eliminates the need for a prior knowledge of the total fertility rate to calculate the proportions.

The information pertaining to each age group of mothers generates an estimate of the total fertility rate. It is recommended to analyze the estimated levels and their variations by age. Frequently, the estimates pertaining to ages 20 to 24 years and 25 to 29 years are taken as the actual levels of the total fertility rate.

The Bureau of the Census has developed the spreadsheet REL-GMPZ to calculate the estimates by this technique. The estimates are derived by using both the average number of children ever born by age of mother and the fertility rates based on children born during last year. See appendix IV-8, and documentation in volume II.
Fertility rates pertaining to a population are related to one another. A total fertility rate should be consistent with the crude birth rate and the general fertility rate. The total fertility rate is derived from a set of age-specific fertility rates which, when applied to the female population, produce a total number of births and hence a crude birth rate. Therefore, if age-specific fertility rates, female population by 5-year age groups, and the total population for both sexes combined are available, it is possible to calculate consistent values of the total fertility rate, general fertility rate, and crude birth rate.

The need for consistency described above was the motive for developing four spreadsheets which, under certain assumptions, can be used to estimate selected indices based on fertility and population information. These four spreadsheets are: ADJASFR, TFR-GFR, CBR-GFR, and CBR-TFR.

The last three spreadsheets use patterns of age-specific fertility rates obtained from empirical observations pertaining to developing countries. The empirical sets of age-specific fertility rates were grouped in clusters. To pertain to a cluster, total fertility rates had to be close to integer units. Eight clusters were made, for total fertility rates close to 1, 2, 3, ..., 7 and 8. The age-specific fertility rates within each cluster were averaged and adjusted to total fertility rates of exactly 1, 2, 3, ..., 7 and 8.

To select the fertility rates used in the programs, each spreadsheet interpolates between two consecutive sets of age-specific fertility rates to select the fertility rates used in the program. If any of the three spreadsheets has to use a set of age-specific fertility rates pertaining to a total fertility rate under 1 or above 8, the spreadsheet will indicate errors.

**Spreadsheet ADJASFR**

**Description**

This spreadsheet adjusts a given pattern of age-specific fertility rates to reproduce a desired total number of births, given the female population in childbearing ages, by 5-year age groups, and a total population for both sexes combined.

**Data Required**

1. Female population in childbearing ages, by 5-year age groups.
2. Total population for both sexes combined.
A pattern of age-specific fertility rates.

A desired total number of births.

Procedure

Given the female population and a pattern of fertility rates, the number of births implied by both sets of data is calculated. This total number of births is compared with the desired total number of births. The pattern of age-specific fertility rates is proportionally adjusted to reproduce the desired total number of births. With the adjusted age-specific fertility rates, the total fertility rate and crude birth rate are calculated, corresponding to the desired number of births and the total population for both sexes combined.

Software

Documentation of the Bureau of the Census spreadsheet ADJASFR is presented in volume II.

Spreadsheet TFR-GFR

Description

This spreadsheet estimates the total fertility rate and the general fertility rate based on a given female population in childbearing ages, by 5-year age groups, a crude birth rate, the total population for both sexes combined, and sets of age-specific fertility rates empirically derived and provided in the spreadsheet.

Data Required

1. Female population, by 5-year age groups in childbearing ages.

2. Total population for both sexes combined.

3. A desired crude birth rate.

Procedure

The program contains eight empirical sets of age-specific fertility rates for levels of total fertility rates from 1 to 8, with unit increments. With these sets of rates, the following procedure is used:
program which are applied to the female population in childbearing ages. The births obtained are used to calculate crude birth rates, since the total population for both sexes is also given as input.

(2) The program searches for the two crude birth rates calculated in step (1) that surround the desired birth rate. Once the two birth rates are selected, they are used, together with the desired rate, to interpolate between the two sets of age-specific fertility rates.

(3) The interpolated fertility rates from step (2) are used to calculate a total fertility rate and general fertility rate consistent with the desired crude birth rate.

If the interpolated set of age specific fertility rates pertain to a total fertility rate under 1 or above 8, the spreadsheet will indicate errors.

Software

Documentation of the Bureau of the Census spreadsheet TFR-GFR is presented in volume II.

**Spreadsheet CBR-GFR**

**Description**

This program calculates the crude birth rate and the general fertility rate based on a desired total fertility rate, the female population in childbearing ages, by 5-year age groups, the total population for both sexes combined, and sets of age-specific fertility rates derived empirically and provided in the spreadsheet.

**Data Required**

(1) Female population in childbearing ages, by 5-year age groups.

(2) Total population of both sexes combined.

(3) A desired total fertility rate.
Procedure

The program contains eight empirical sets of age-specific fertility rates for levels of total fertility rates from 1 to 8, with unit increments. With these sets of rates, the following procedure is used:

1. There are eight sets of age-specific fertility rates with total fertility rates of 1, 2, ..., and 8. The program searches for the two total fertility rates that surround the desired total fertility rate. The two selected rates and the desired one are used to interpolate the age-specific fertility rates.

2. The interpolated age-specific fertility rates from step (1) are used, together with the information on females in childbearing ages and total population for both sexes, to calculate the total number of births, the crude birth rate, and the general fertility rate.

If the interpolated set of age-specific fertility rates pertains to a total fertility rate under 1 or above 8, the spreadsheet will indicate errors.

Software

Documentation of the Bureau of the Census spreadsheet CBR-GFR is presented in volume II.

Spreadsheet CBR-TFR

Description

This program estimates the crude birth rate and the total fertility rate, based on the female population in childbearing ages, by 5-year age groups, the total population for both sexes combined, the general fertility rate, and sets of age-specific fertility rates derived empirically and provided in the spreadsheet.

Data Required

1. Female population in childbearing ages, by 5-year age groups.

2. Total population for both sexes combined.

3. A desired general fertility rate.
Procedure

The program contains eight empirical sets of age-specific fertility rates for levels of total fertility rates from 1 to 8, with unit increments. With these sets of rates, the following procedure is used:

(1) There are eight sets of age-specific fertility rates (with total fertility rates of 1, 2, ..., and 8) provided by the program which applies to the female population in childbearing ages. The births obtained are used to calculate the general fertility rates.

(2) The program searches for the two general fertility rates calculated in step (1) that surround the desired general fertility rate. Once the two rates are selected, they are used, together with the desired rate, to interpolate between the two sets of age-specific fertility rates.

(3) The interpolated set of age-specific fertility rates from step (2) is used to calculate the number of births, the crude birth rate, and the total fertility rate.

If the interpolated set of age-specific fertility rates pertains to a total fertility rate under 1 or above 8, the spreadsheet will indicate errors.

Software

Documentation of the Bureau of the Census spreadsheet CBR-TFR is presented in volume II.
Appendix IV-2

Standardization of the Total Fertility Rate

In order to standardize the total fertility rate, the age-specific fertility rates have to be decomposed into two or more components. In this exercise, we will decompose the fertility rate into the more frequently used components: the proportion of women who are mothers and the fertility of mothers.

Sometimes the proportion of women ever married is used instead of the proportion of mothers. As discussed in the text, the proportion of women ever married is not the same as the proportion of women who had or are having children, but it is a good proxy. In most countries, the proportion of women who are mothers can be determined from census tabulations which also present the total number of children that women have had.

The decomposition is accomplished by the following formulas:

$$5f_x = \frac{5B_x}{5FP_x} = \frac{5B_x}{5W_x} \cdot \frac{5W_x}{5FP_x} = 5wf_x \cdot 5pw_x$$

Where:

- $5f_x$ represents the age-specific fertility rate pertaining to all women age $x$ to $x+4$ years;
- $5B_x$ represents the number of births to mothers age $x$ to $x+4$ years;
- $5FP_x$ represents the total number of women age $x$ to $x+4$ years;
- $5W_x$ represents the number of women who are mothers at age $x$ to $x+4$ years;
- $5wf_x$ is the fertility rate of all mothers in the population at age $x$ to $x+4$ years; and
- $5pw_x$ is the proportion of mothers age $x$ to $x+4$ years.

Hence, the total fertility rate can be expressed in terms of mother fertility and proportion of mothers, as follows:

$$TFR = 5 \sum 5f_x = 5 \sum 5wf_x \cdot 5pw_x$$

Where the symbols are the same as above.
If there is a population with information as presented in table A-IV-2.1, the following products can be calculated:

\[ K = \sum x \cdot f_x \cdot p_x(2) \]  
\[ Q = \sum x \cdot f_x \cdot p_x(1) \]

Where the numbers in parentheses indicate the first and second years of information.

Taking into account that the total fertility rate for each year is the sum of such products as calculated for each age group using the fertility and proportions for the same year, the following components of the differences between total fertility rates for the two years can be determined:

1. Difference due to a change in the proportion of females who are mothers:
\[ D_M = TFR(1) - K \]

2. Difference due to a change in the fertility of mothers:
\[ D_F = TFR(1) - Q \]

3. Interaction:
\[ I = TFR(1) - TFR(2) - K - Q \]

Where the symbols are the same as above.

If the total change in the total fertility rates is \( T_D = TFR(1) - TFR(2) \), then the percent contribution of each factor (mother fertility, proportion of mothers, and interaction) is, respectively:

\[
\begin{array}{ccc}
DM & DF & I \\
100 & 100 & 100 \\
TD & TD & TD \\
\end{array}
\]

In the example presented in table A-IV-2.1, it can be concluded that of the reduction of the total fertility rate of 0.6 children per woman, 48.6 percent was due to the changing proportion of women who have children, and 51.5 percent to the fertility of women having children.
Table A-IV-2.1. Standardization of Total Fertility Rates by Proportion of Married Females and Marital Fertility

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Females</th>
<th>Births</th>
<th>Married Females</th>
<th>Married</th>
<th>Married Females</th>
<th>Married</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>118609</td>
<td>11624</td>
<td>35723</td>
<td>137264</td>
<td>12043</td>
<td>36149</td>
</tr>
<tr>
<td>20-24</td>
<td>103252</td>
<td>22143</td>
<td>69069</td>
<td>119954</td>
<td>23034</td>
<td>71634</td>
</tr>
<tr>
<td>25-29</td>
<td>85086</td>
<td>23432</td>
<td>76573</td>
<td>90365</td>
<td>22623</td>
<td>76599</td>
</tr>
<tr>
<td>30-34</td>
<td>69512</td>
<td>15678</td>
<td>64160</td>
<td>73645</td>
<td>15743</td>
<td>67345</td>
</tr>
<tr>
<td>35-39</td>
<td>57470</td>
<td>10834</td>
<td>53217</td>
<td>62834</td>
<td>10865</td>
<td>58024</td>
</tr>
<tr>
<td>40-44</td>
<td>46958</td>
<td>4555</td>
<td>43765</td>
<td>54876</td>
<td>4265</td>
<td>51534</td>
</tr>
<tr>
<td>45-49</td>
<td>38435</td>
<td>1507</td>
<td>35821</td>
<td>42765</td>
<td>1243</td>
<td>39704</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age Group</th>
<th>All Married</th>
<th>Multiplication of columns (2)(3)(5)</th>
<th>All Married</th>
<th>Multiplication of columns (2)(3)(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>.0980</td>
<td>.3254</td>
<td>.3012</td>
<td>.0877</td>
</tr>
<tr>
<td>20-24</td>
<td>.2164</td>
<td>.3206</td>
<td>.6750</td>
<td>.1922</td>
</tr>
<tr>
<td>25-29</td>
<td>.2754</td>
<td>.3068</td>
<td>.8976</td>
<td>.2504</td>
</tr>
<tr>
<td>30-34</td>
<td>.2255</td>
<td>.2444</td>
<td>.9230</td>
<td>.2138</td>
</tr>
<tr>
<td>35-39</td>
<td>.1885</td>
<td>.2036</td>
<td>.9260</td>
<td>.1726</td>
</tr>
<tr>
<td>40-44</td>
<td>.0970</td>
<td>.1041</td>
<td>.9320</td>
<td>.0774</td>
</tr>
<tr>
<td>45-49</td>
<td>.0340</td>
<td>.0365</td>
<td>.9320</td>
<td>.0291</td>
</tr>
<tr>
<td>TFR</td>
<td>5.674</td>
<td></td>
<td>5.116</td>
<td></td>
</tr>
</tbody>
</table>

Special calculations:

Decomposition of the difference in the total fertility rate Percent

| Age Group | Multiplication of columns (2)(3)(5) | Total (5.674-5.118) | .559 | 100.0
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>.0857</td>
<td></td>
<td>.1003</td>
<td></td>
</tr>
<tr>
<td>20-24</td>
<td>.1916</td>
<td></td>
<td>.2100</td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>.2601</td>
<td></td>
<td>.2651</td>
<td></td>
</tr>
<tr>
<td>30-34</td>
<td>.2235</td>
<td></td>
<td>.2158</td>
<td></td>
</tr>
<tr>
<td>35-39</td>
<td>.1880</td>
<td></td>
<td>.1731</td>
<td></td>
</tr>
<tr>
<td>40-44</td>
<td>.0977</td>
<td></td>
<td>.0768</td>
<td></td>
</tr>
<tr>
<td>45-49</td>
<td>.0339</td>
<td></td>
<td>.0292</td>
<td></td>
</tr>
<tr>
<td>TFR</td>
<td>5.402</td>
<td></td>
<td>5.386</td>
<td></td>
</tr>
</tbody>
</table>

Note: TFR is the total fertility rate.

Source: U.S. Bureau of the Census.
Appendix IV-3

Rejuvenation Technique

Description

The rejuvenation technique uses a current population, by age and sex, to estimate the number of persons, by age and sex, during the past, including an estimate of the number of births for specific years. Based on the estimated number of births and population, crude birth rates for 5-year periods are calculated.

Data Required

1. Male and female age distributions of the population, by 5-year age groups.
2. Male and female life table \( l_x \) values for two dates (preferably surrounding the rejuvenation period).

Assumptions

1. The completeness of enumeration of the population is the same in all ages.
2. There is no age misreporting.

Procedure

The population in each age group is divided by the corresponding survival ratio to rejuvenate the population 5 years. The age groups under age 1 year, 1 to 4 years, 5 to 9 years, and 10 to 14 years are used to estimate the number of births during the last year and for periods 1 to 4 years, 5 to 9 years, and 10 to 14 years prior to the census or survey date, respectively. The total population is estimated for the midyear of those intervals, based on the input growth rates.

Advantages

It is a simple procedure that, with good information, can provide acceptable estimates of the birth rate for comparison with rates estimated using other procedures.
Limitations

(1) A higher undercount of the young population (particularly under age 5 years) than in other ages will produce a crude birth rate lower than the actual rate.

(2) For countries with a rather recent reduction in fertility, it may not be possible to determine whether a decline in the estimated crude birth rate during the 5 years prior to the census date is real or is attributable to an undercount of the population under age 5 years.

(3) The estimated crude birth rates pertain not to the current year but to selected periods prior to the census.

Software

There is a spreadsheet prepared by the Bureau of the Census to apply this procedure, called REVCBR. Its documentation is presented in volume II.
Appendix IV-4

Own-Children Technique

Description

This technique estimates age-specific fertility rates based on a current age structure of the population and mortality estimates for the past. The time reference of the estimates runs from the year prior to the census to about 15 years back in time.

Data Required

(1) A distribution of the population by single years of age.
(2) Mortality estimates for a 15-year period prior to the census.
(3) Census information that matches each child to its mother.

Assumptions

(1) Completeness of census enumeration is the same for all ages.
(2) There is no age misreporting.
(3) There is no migration.

Procedure

The technique requires special tabulations matching children with their natural mothers. Once children and mothers are matched, the number of children and the female population are rejuvenated in such a way that children are continuously linked to their mothers, by age. This procedure allows the estimation of the number of births, by age of mother, for each year. Based on the number of births by age of mother and the female population, age-specific fertility rates are calculated. In cases where information is not available to match children with their mothers, a simplified version of this technique can be applied, but the results are less reliable.

Advantages

The technique provides estimates of the total fertility rate as well as age-specific fertility rates without requiring any assumptions concerning the pattern of fertility.
Limitations

Fertility estimates are based on the assumption that completeness of enumeration was the same for all ages. However, in developing countries, the completeness of enumeration for ages under 5 years is usually lower than for other age groups. This produces biased estimates, since the number of births estimated based on the youngest ages is too small. As a result, the trend of the fertility estimates shows a decline for 2 or 3 years prior to the census, due to the higher undercount of children ages under 1 year, 1 year, and 2 years.
Appendix IV-5

Rele Technique

**Description**

This technique estimates the gross reproduction rate (and hence the total fertility rate, if the sex ratio at birth is available) of a population for one or two 5-year periods prior to the census date. The estimates are based on the number of children under age 10, the number of women in reproductive ages, and the life expectancy at birth.

**Data Required**

1. The child population (both sexes combined) ages 0 to 4 years and 5 to 9 years.
2. The female population ages 15 to 44 years, 15 to 49 years, 20 to 49 years, and 20 to 54 years.
3. The life expectancy at birth for both sexes combined.

**Assumptions**

1. The population is stable or quasi-stable.
2. There is no migration.
3. Completeness of enumeration is the same in all the required age groups.
4. Age reporting of children is correct within each 5-year age group.
5. Age misreporting of women within the groups of childbearing ages does not affect the total number of women in those ages.
6. The mean length of generation is 28 to 29 years.

**Procedure**

Rele observed that in stable populations, at a given level of mortality, there is an almost linear relationship between the child/woman ratio and the fertility level. Based on this observation, he correlated the gross reproduction rate with such ratios for different levels of life expectancy at birth. Hence, given a life expectancy at birth and child/woman ratios, an estimate of the gross reproduction rate can be obtained by using the correlation coefficients. Although the correlations made by Rele provide
estimates of only the gross reproduction rate, total fertility rates can be calculated from them if the sex ratio at birth is known.

Advantages

(1) The technique requires little information. All modern censuses provide information on the population by age and sex.

(2) The technique is not particularly sensitive to the level of mortality, and so errors in estimating the life expectancy at birth have little effect on the estimate of the gross reproduction rate.

Limitations

(1) Differential undercount in the age groups used will produce biased results. Frequently, a higher undercount of the population ages 0 to 4 years will produce a declining trend of fertility that may be incorrect.

(2) An undercount of the population under age 5 years may be mistaken for a recent decline in fertility.

(3) The departure of the population from stable conditions will produce biased estimates. Changing fertility and a high level of migration will have a greater impact than changing mortality.

(4) Strong age misreporting will affect the results.

Note

In order to estimate the sensitivity of this technique to the levels of mortality and sex ratios at birth, the technique can be applied to the same population, but with different combinations of life expectancies at birth and sex ratios at birth.

Software

The spreadsheet RELEFERT developed by the Bureau of the Census calculates the estimates. Its documentation is presented in volume II.
Appendix IV-6

P/F Ratio Technique

Description

The P/F ratio technique, originally developed by William Brass, produces a factor for adjusting reported age-specific fertility rates (based on vital registration or births in the 12 months prior to a census or survey) to the "actual" level of fertility. The adjustment factor is determined by comparing data on children ever born, by age of women, with a set of cumulated age-specific fertility rates.

Data Required

(1) Average number of children ever born alive, per woman, by 5-year age groups (15 to 19 years to 45 to 49 years), as collected in a census or survey.

(2) Age-specific fertility rates for the same 5-year age groups of women. These rates may be based on reported births in the 12 months prior to the census or on independent registration for a time period close to the date of the census or survey.

(3) Information to determine whether the age of women reporting births in the last 12 months was reported as of the date of the childbirth or as of the date of the census or survey.

Assumptions

(1) The reporting of the average number of children ever born is complete (at least for younger women, under 30 years or 35 years of age), and represents the level of cumulative fertility up to these ages.

(2) The completeness of reporting of births used to estimate the age-specific fertility rates is the same for all age groups of women.

(3) The pattern and level of fertility have not changed in the recent past (15 to 20 years prior to the census or survey).

Procedure

The basis of this technique is a comparison of the average number of children ever born by age of women (symbolized by P) with cumulative age-specific fertility rates from registers or from reported births during the last year (symbolized by F). This comparison requires some adjustments, as
discussed below. After such adjustments, the ratios P/F produce a factor for adjusting the age-specific fertility rates to the "actual" level of fertility.

Although the concepts of cumulative fertility rates and average number of children ever born per woman are similar (both represent cumulative fertility), the age of the cumulation is not the same. The average number of children ever born pertains to the mean age of each interval (17.5 years, 22.5 years, ... , 47.5 years), while the cumulative fertility rates pertain to the end of each age interval (20 years, 25 years, ... , 50 years). In other words, there are 2.5 years difference between them. Even if the information were perfect, this difference would produce higher values for cumulative fertility than for the average number of children ever born for each age group.

Three methods have been developed to adjust the data for this age discrepancy. Brass, the author of the technique, simulated the average number of children ever born per woman and the cumulative fertility by using a third degree polynomial. Later, Trussell and Coale developed two alternatives. One of them uses second degree polynomials for each three consecutive observations. The other uses the least square technique and model age-specific fertility rates to obtain a set of coefficients for the adjustment.

An additional adjustment is sometimes required. For cases in which the age of the women was reported as of the census or survey date, the age-specific fertility rates based on births during the last year occurred when the women were, on average, one-half year younger. This requires that the pattern of the age-specific fertility rates be adjusted by one-half year.

Once the two adjustments are made, the ratios of P to F can be taken and applied to the age-specific fertility rates to proportionally increase or decrease them to the "proper" level of fertility represented by the average number of children per woman at young adult ages. The ages most frequently used are 20 to 24 years and 25 to 29 years; occasionally, 30 to 34 years.

The analysis of the P/F ratios is important. For instance, if the data are accurate and fertility has been decreasing in recent years, the P/F ratios may show a trend that rises with the age groups of women. This is because most of the births to older women occurred in the past when fertility was higher, while births to younger women occurred more recently when fertility had already declined. This situation occurs especially when a rise in the age at marriage leads to a reduction in fertility among the youngest age groups of women. Under these conditions, the results of the technique do not represent the current pattern of cumulated fertility.

Advantages

(1) The method is not affected by the failure of some women to report the number of children ever born if the women who did not report have the same number of children as those who reported.

(2) The method uses a relatively small amount of information, although it requires special questions in the survey or census.
Limitations

(1) If fertility has not been constant, the results of the technique may be biased upward.

(2) Errors in the data on children ever born will affect the results:

(a) Age misreporting of women providing these data will have an unpredictable effect.

(b) Underreporting of children ever born will cause a downward bias in the adjusted estimates. Children who died in infancy (especially in very early infancy), as well as those living away from home, are the births most likely to be omitted, especially by older women.

(c) Overreporting of children ever born will cause an upward bias in the adjusted estimates. Overreporting of children can sometimes occur when stillbirths, late foetal deaths, or adopted children are mistakenly included.

(3) Errors in the age-specific fertility rates will affect the results:

(a) Age misreporting of women providing these data will have an unpredictable effect.

(b) If the pattern of fertility taken as the "actual" pattern contains errors, the estimated age-specific fertility rates will be incorrect. This may also affect the level of the total fertility rate.

Software

The Bureau of the Census has developed a spreadsheet to apply this technique. It is called PFRATIO and its documentation is presented in volume II.
Appendix IV-7

Arriaga Technique

Description

This technique estimates fertility rates based on information on the average number of children ever born, by age of mother, collected in censuses or surveys and, if available, a pattern of fertility based on vital registration or on the number of births during the 12 months prior to the same censuses or surveys. The following discussion applies to the situation where data on children ever born and age patterns of fertility are available for at least two dates. If data are available for only one date, then the assumptions, advantages, and limitations are basically the same as those for the P/F ratio technique.

Data Required

(1) The average number of children ever born per woman, by 5-year age groups, for two dates.

(2) An age pattern of fertility, that is, a distribution of age-specific fertility rates by 5-year age groups. Although vital statistics can provide a pattern of fertility, frequently the pattern is obtained from information on the number of births occurring during the 12 months prior to a census or survey, classified by age of mother. Although the estimates can be obtained even if a pattern of fertility is not available, better results are achieved if such information is used.

Assumptions

(1) The completeness of reporting of births used to estimate the age-specific fertility rates is the same for all age groups of women.

(2) Reporting of the average number of children ever born per woman is complete (at least for women under 30 years or 35 years of age).

(3) Changes in fertility produce a linear change in the average number of children ever born per woman at each particular age of woman (mainly at ages 15 to 35 years) between the two dates.

(4) Fertility occurs only between exact ages 15 and 50 years.

Procedure

(1) The average number of children ever born per woman is estimated by single ages of the woman for the following dates: (a) the dates
of the censuses or surveys; (b) 1 year before the latest census or survey date; (c) 1 year after the earliest census or survey date; and (d) 1 year before and after any intermediate census or survey dates. The estimates for census or survey dates are obtained by an interpolation process with a nine degree polynomial. For years before or after the censuses, the estimates are obtained by linear interpolation using the single-age estimates from censuses or surveys.

(2) Age-specific fertility rates for single ages are calculated based on annual cohort changes in children ever born using the estimates obtained in (1).

(3) Fertility rates for 5-year age groups are calculated by averaging the single-age rates obtained in (2) within the age group.

(4) Cumulative fertility rates are calculated for the rates obtained in (3) and for the rates pertaining to the pattern of fertility.

(5) Adjustment factors are calculated by dividing the cumulated fertility rates originated by comparing the increase in the number of children ever born in each cohort as calculated in (3) by the corresponding cumulated fertility rates pertaining to the pattern of fertility.

(6) The final age-specific fertility rates are estimated by multiplying the rates from the pattern of fertility by a selected adjustment factor as calculated in (5). It is recommended to select the adjustment factor that corresponds to the age group whose mean is closest to the mean age of the fertility pattern.

Advantages

(1) Since the technique does not assume that fertility is constant, it can provide an estimate of fertility when it has been changing.

(2) Fertility estimates are obtained for the year of the censuses or surveys.

(3) An analysis of the adjustment factors allows for an evaluation of the data used. The adjustment factors are not affected by changing fertility but are affected by the quality and comparability of the data.

Limitations

(1) Errors in the data on children ever born will affect the results:

   (a) Age misreporting of women providing these data will have an unpredictable effect.
(b) Underreporting or overreporting of children ever born by women under age 35 years will affect the estimates. If the reporting of children ever born is more (or less) complete in one census than in the other, the estimates of fertility will be affected, and the trend in fertility may be incorrect. This can occur when data from censuses and surveys are combined in the estimation process because survey data are frequently of better quality than census data.

(c) If the technique is used when a pattern of fertility is not available, the results should be interpreted with caution. It is advisable to make graphs of the single-year age-specific fertility rates to evaluate the results.

(2) Errors in the age-specific fertility rates will affect the results:

(a) Age misreporting of women providing these data will have an unpredictable effect.

(b) If the pattern of fertility taken as the "actual" pattern contains errors, the estimated age-specific fertility rates will be incorrect. This may also affect the level of the total fertility rate.

Software

There are several computer programs to apply this technique. The Bureau of the Census has developed two spreadsheets: ARFE-2 and ARFE-3. The first one is used when information on children ever born and an age pattern of fertility are available for two dates, and the second one is used when such information is available for three dates. Documentation of these programs is presented in volume II. The United Nations also has developed two programs in the MORTPAK package: FERTCB and FERTPF. The first one estimates fertility without using the pattern of fertility, while the second one is similar to ARFE-2.
Appendix IV-8

Relational Gompertz Technique

Description

This technique was developed by Brass and associates "for the evaluation and adjustment of fertility estimates obtained from retrospective reports of birth histories or features of birth histories." The technique can also be used to estimate total fertility rates. The estimates are based on information on the average number of children ever born per woman, by 5-year age groups of women in childbearing ages. If age-specific fertility rates based on children born during the year prior to the census or survey are also available, additional estimates can be obtained.

Data Required

(1) Average number of children ever born per woman, by 5-year age groups of mothers in childbearing ages.

(2) Optional: Age-specific fertility rates based on information on children born during the year prior to the census or survey date.

Assumptions

(1) The average number of children ever born per woman, by age of the women, follows the pattern of a Gompertz function.

(2) The reporting of the average number of children ever born per woman, by age of women, is complete and represents the level of cumulative fertility up to each age group.

(3) The completeness of reporting of children born during last 12 months prior to the census or survey is the same for all age groups of women.

Procedure

The technique fits a Gompertz function to the information on the average number of children ever born, by age of the woman. The function is:

\[ F(x) = F \cdot A^x \]

Where:

\( F(x) \) represents the cumulative fertility up to age \( x \), or the average number of children per woman by age of the woman from census or survey information;
F is the total fertility rate;
A and B are constants; and
x represents age.

The author transforms the double exponential function into a linear one by taking logarithms, twice. In addition, a scale transformation is performed to obtain a better fit of the Gompertz function to the actual data. The author uses standards for adjusting the actual data to the shape of the Gompertz function.

The problem is that F, the total fertility rate, has to be known for the linearization of the Gompertz function. This problem was solved by Zaba (1981) using Taylor series. Finally, the technique uses one standard and a set of coefficients for estimates derived from information on children ever born per woman, by age of the woman. Another standard and set of coefficients are used if estimates of total fertility rates are desired using the pattern of fertility from information on children born during the 12 months prior to the census or survey. While the technique does not require a pattern of fertility (it is optional), it does require the average number of children ever born per woman, by age of mother. This is required even if the estimation will be based on children born during the last 12 months.

Advantages

The technique provides an estimated total fertility rate based on each 5-year age group in childbearing ages. An analysis of the fluctuation of the estimates may be used to infer a trend in the fertility level.

Limitations

(1) Errors in the average number of children ever born per woman, by age of the woman, affect the results.

(2) Estimates based on information for the age group 15 to 19 years are sensitive to information errors. Such estimates should be taken with caution.

(3) Estimates based on data on children born during the year prior to the census or survey fluctuate more than those based on the average number of children ever born. While this could make difficult the selection of a level, it is also an advantage in that it reveals inconsistencies between the average number of children ever born per woman and the number of children born during the last year.
Software

The Bureau of the Census has developed a spreadsheet called REL-GMPZ, which performs the calculations of this technique. Its documentation is presented in volume II.
REFERENCES


Chapter V

MIGRATION

This Chapter in Brief

Since few countries in the world have population registers to collect information on migration, this component of population change has to be estimated based on census and/or survey information. Migration is divided into internal and international migration.

Internal migration

The techniques for detecting the number of migrants are based on two kinds of information:

(1) Information collected specifically for the measurement of migration, such as place of previous residence.

(2) Information not collected specifically for the purpose of measuring migration, such as the age structure of a population. See Bureau of the Census spreadsheet CSRMIG.

International migration

Techniques are based on information on place of birth and other questions for detecting out-migration.

Spreadsheet and Method That Can Be Used for Analyzing Population Migration, According to the Available Information

<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age specific population data</td>
<td>CSRMIG</td>
<td>Estimates net migration in an area.</td>
</tr>
</tbody>
</table>
Where to Find the Software

The following spreadsheet for measuring migration of a population can be found in volume II.

Introduction

With the improvement of economic conditions and the increase of communication and transportation systems, people increase their desire to change residence. A mere change of residence, however does not always constitute a migratory movement. Although there is no precise definition for migration, it is understood that it involves a certain distance. Thus, a change of residence within a relatively small area (a city or the smallest administrative division) is not considered a migratory movement. A migratory move also implies an intention that the move be permanent. Hence, a migrant is a person who moves a certain distance with the intention that the move be permanent, and the move affects the population growth of the areas of both origin and destination. Migration is, therefore, the third component of the population growth of an area, together with fertility and mortality.

From an administrative or legal point of view, migration may be classified into two categories: internal and international. Internal migration is the movement within the boundaries of a single country, while international migration is the movement from one country to another.

Although migration is sometimes an important component of population growth, it is not well recorded in most countries of the world. Only a few developed countries maintain population registers which record not only births and deaths but also other changes in the status of its citizens, such as civil status, employment, and, of course, change of residence. The Netherlands has such a system, and migration for that country is known with an acceptable degree of accuracy. But the Netherlands is an exception; in most countries, migration is estimated indirectly, resulting in only a partial detection of the total movement. If the population of a country or of an area within a country is growing at a rate different from the rate of natural increase, the difference is due to migration.
This chapter presents some of the procedures most frequently used for measuring migration: 
(a) procedures based on questions in censuses or surveys intended for detecting migration; and 
(b) procedures based on other population characteristics, such as age structure and place of birth. Internal migration is presented first, and then international migration is discussed.

**Internal Migration**

Once economic development has begun in a country, internal migration becomes an important component of population growth, especially in those areas attracting the migrants. Job opportunities, better salaries, educational possibilities, and other factors start to attract people from other areas of the country. To plan for the redistribution of the population, it is necessary to know the number and characteristics of the migrants. There are not many procedures for measuring migration. The following sections present the more frequently used methods, given the information that is usually available.

**Procedures based on questions for detecting migration**

Migration requires a change of residence, a certain distance from previous to current residence, and a time period of reference. The latter is required in order to estimate the number of migrants per year or for any other specified period. Based on this general definition of migrants, questions are often included in censuses and surveys about the previous residence of the population enumerated in a given place. The respondent may be asked either to state the year of change of residence or to specify whether it took place during a given period of reference, usually about 5 years. In the latter case, the information collected is the place of residence 5 years ago. Cross-tabulations of the population by place of enumeration and place of previous residence provide a measure of migration.

**Place of enumeration and place of previous residence.** When the period of reference is 5 years, migration is estimated for the 5-year period prior to the census or survey. If the census asked the year of movement, other periods of reference could be obtained. However, experience has shown that responses to a question on the year of migration are usually unreliable, and so the question about residence 5 years ago is the most common. Although the procedure described below is based on the assumption that the data refer to a 5-year period, it would be similar for a different reference period.

Responses to the question on residence 5 years ago may be used to produce a tabulation which gives a general idea of the migration movement in a country. An example of such a tabulation is presented in table V-1. Since the period of reference is 5 years prior to the census date, the population under age 5 is not included. Each column represents the place of residence 5 years ago, and each row represents the place of residence at the time of enumeration. The diagonal from top left to bottom right represents persons who were enumerated in the same place as their residence 5 years ago. This does not necessarily mean that they did not migrate, however. Some persons
enumerated in state B whose residence 5 years ago was also state B (a total of 6,821 in table V-1) could have moved to another state 4 years ago and more recently (perhaps 1 year ago) returned to state B. This kind of migration movement would not be detected by the question on residence 5 years ago. For this reason, persons who were enumerated in a state different from their residence 5 years ago will be referred to as "detectable migrants," and those enumerated in the same state will be referred to as "apparent nonmovers."

### Table V-1. Internal Migration Based on Questions on Previous Residence

<table>
<thead>
<tr>
<th>State of enumeration</th>
<th>State of residence 5 years prior to census date</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>2,518</td>
<td>152</td>
</tr>
<tr>
<td>B</td>
<td>58</td>
<td>6,821</td>
</tr>
<tr>
<td>C</td>
<td>105</td>
<td>99</td>
</tr>
<tr>
<td>D</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>E</td>
<td>41</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>2,754</td>
<td>7,120</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of the Census.

This cross-tabulation has the advantage of providing data about origin and destination of migratory movements. For example, it is known that 7,530 persons were enumerated in state C (total of the row pertaining to state C), while only 7,314 persons were living in that state 5 years ago (those who died and children under age 5 are not taken into account). These figures indicate a net migration to state C of 216 persons (7,530 - 7,314). The total number of detectable in-migrants is calculated as the total number of persons enumerated in state C minus the apparent nonmovers (those who were enumerated in state C and also lived there 5 years ago); in the table, this group is 7,530 - 7,138 = 392. The detectable migrants came from the following states, A = 105, B = 99, D = 68, and E = 120 (table V-1, row pertaining to state C). Similarly, the number living in state C 5 years ago (regardless of where they were enumerated now) minus the number of apparent nonmovers gives the detectable out-migration from state C (7,314 - 7,138 = 176). Those who left went to the following states: A = 73, B = 69, D = 17, and E = 17 (table V-1, column pertaining to state C).
This tabulation allows the estimation not only of the total migration movement, but of migration rates from each state to each of the other states. Annual migration rates are calculated by dividing the number of a particular group of migrants by the average population of the state of origin during the 5-year period and by the 5 years. For instance, the total annual average out-migration rate per 1,000 persons residing in state C based on the information in table V-1 is:

\[
\frac{[1,000 \times (7,314 - 7,138)]}{1/2 \times (7,530 + 7,314)} / 5 = 4.74
\]

In other words, an average of 4.74 persons for every 1,000 people residing in state C moved out each year during the 5-year period preceding the census.

Similarly, the annual average out-migration rate per 1,000 from state C to state B is:

\[
\frac{[1,000 \times 69]}{1/2 \times (7,530 + 7,314)} / 5 = 1.86
\]

This indicates that an annual average of 1.86 persons for every 1,000 people residing in state C left state C and went to live in state B during the 5-year period preceding the census.

The examples shown in table V-1 refer to the whole population 5 years old and over and do not specify the age structure of the migrants. However, the same cross-tabulation can be made for each desired age and sex group to obtain a detailed knowledge of the migration flow into and out of each state or any other administrative division.

Data collected in censuses concerning place of enumeration and place of residence 5 years ago, particularly when tabulated by sex and age, are probably the best available information for detecting internal migration in countries where population registers do not exist.

Information for children under age 5 can also be obtained if the census includes a question concerning place of birth for those under age 5 and if the data are cross-tabulated by place of enumeration and place of birth. The implicit assumption will be that children under 5 years of age migrated only once since birth.

For areas smaller than states, the same data are usually tabulated but without indication of the place of residence 5 years ago. Instead, the total number of persons who were residing in another place 5 years ago are frequently grouped and presented as in-migrants by sex and age. Presentation of the information in this way does not permit the analysis of out-migration. Thus, cross-tabulations as presented in table V-1 are the best.

**Place of enumeration and place of birth.** This information also permits the analysis of internal migration. Tabulations are made in the same format as in the case of place of enumeration by place of residence 5 years ago. The meaning of the tabulation differs, however, because the reference period is "lifetime" rather than a 5-year period. Not only does the information pertain to an indeterminate period, but the implied flow of migrants from one state to
another could be misleading. Persons who lived in several states during their lifetimes would appear to have made only one migration movement: from the state of birth to the state of current residence. For example, a person born in state A who lived there for several years, then moved to state C for a few years, and subsequently moved to state D, would appear, in the "lifetime" migratory cross-tabulation, to have migrated only from state A to state D.

If more than one census is available with information on place of enumeration by place of birth, a comparison of tabulations from the two censuses may provide some information for the intercensal period, although these estimates will still contain the biases discussed above. For instance, table V-2 presents information from a fictitious country on the population by place of enumeration and place of birth. The first two panels of the table present the total population of the country for the years 1970 and 1980, including persons of foreign origin (tables V-2-A and V-2-B). The differences between the two tabulations, cell by cell, give the minimum number of persons who migrated during the intercensal period (table V-2-C). In this case, the diagonal does not have a meaning with respect to migration, as it includes persons born in each state during the intercensal period and still living in the state; hence the differences have not been calculated. The column presenting the totals of each row represents the number of net in-migrants arriving in each state during the intercensal period (deaths are not considered). For instance, 632 migrants arrived in state B. The negative value (-134) for state A represents a net out-migration of persons born in other states. The row presenting the total of each column represents the net out-migration of persons born in each state. For instance, 893 persons born in state A are living in other states and hence they out-migrated from state A. In the case of state E, the negative figure (-9) indicates that there were 9 persons born in state E and living in other states in 1970 who probably either returned to state E or went out of the country by 1980.

Lifetime migration rates can be calculated in the same way as those based on data on place of enumeration by place of residence 5 years ago, although the meaning is not necessarily the same.

The analysis of data on place of birth is not as reliable as that of data on place of residence 5 years ago. The problem with data on place of birth, in addition to the indeterminate reference period, is that they do not provide information on the exact state of origin of the in-migrants. In general, data on place of birth should be used only when no other information on migration is available.
Table V-2. Migration Analysis Based on Information on Place of Birth and Place of Enumeration

### A) Information for 1970

<table>
<thead>
<tr>
<th>State of enumeration</th>
<th>A)</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>Country</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>23,453</td>
<td>509</td>
<td>98</td>
<td>134</td>
<td>65</td>
<td>12</td>
<td>26,271</td>
</tr>
<tr>
<td>B</td>
<td>365</td>
<td>43,876</td>
<td>456</td>
<td>543</td>
<td>233</td>
<td>32</td>
<td>45,506</td>
</tr>
<tr>
<td>C</td>
<td>760</td>
<td>476</td>
<td>35,212</td>
<td>234</td>
<td>146</td>
<td>45</td>
<td>36,881</td>
</tr>
<tr>
<td>D</td>
<td>123</td>
<td>212</td>
<td>205</td>
<td>12,011</td>
<td>287</td>
<td>39</td>
<td>12,877</td>
</tr>
<tr>
<td>E</td>
<td>548</td>
<td>876</td>
<td>675</td>
<td>768</td>
<td>67,390</td>
<td>132</td>
<td>70,389</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>25,257</td>
<td>45,949</td>
<td>36,646</td>
<td>13,690</td>
<td>68,121</td>
<td>260</td>
<td>189,923</td>
</tr>
</tbody>
</table>

### B) Information for 1980

<table>
<thead>
<tr>
<th>State of birth</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>Country</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>31,578</td>
<td>402</td>
<td>137</td>
<td>95</td>
<td>43</td>
<td>7</td>
<td>32,262</td>
</tr>
<tr>
<td>B</td>
<td>432</td>
<td>53,751</td>
<td>674</td>
<td>845</td>
<td>276</td>
<td>34</td>
<td>56,012</td>
</tr>
<tr>
<td>C</td>
<td>984</td>
<td>1,678</td>
<td>45,231</td>
<td>528</td>
<td>187</td>
<td>53</td>
<td>48,661</td>
</tr>
<tr>
<td>D</td>
<td>325</td>
<td>879</td>
<td>102</td>
<td>14,973</td>
<td>216</td>
<td>31</td>
<td>16,526</td>
</tr>
<tr>
<td>E</td>
<td>956</td>
<td>2,376</td>
<td>1,376</td>
<td>974</td>
<td>87,465</td>
<td>189</td>
<td>93,336</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>34,275</td>
<td>59,086</td>
<td>47,520</td>
<td>17,415</td>
<td>88,187</td>
<td>314</td>
<td>246,797</td>
</tr>
</tbody>
</table>

### C) Difference between the information for 1980 and 1970, except diagonal

<table>
<thead>
<tr>
<th>State of birth</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>Country</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>-107</td>
<td>39</td>
<td>-39</td>
<td>-22</td>
<td>-5</td>
<td>-134</td>
</tr>
<tr>
<td>B</td>
<td>67</td>
<td>9,875</td>
<td>218</td>
<td>302</td>
<td>43</td>
<td>2</td>
<td>632</td>
</tr>
<tr>
<td>C</td>
<td>216</td>
<td>1,202</td>
<td>0</td>
<td>294</td>
<td>41</td>
<td>8</td>
<td>1,761</td>
</tr>
<tr>
<td>D</td>
<td>202</td>
<td>667</td>
<td>-103</td>
<td>0</td>
<td>-71</td>
<td>-8</td>
<td>687</td>
</tr>
<tr>
<td>E</td>
<td>408</td>
<td>1,500</td>
<td>701</td>
<td>206</td>
<td>0</td>
<td>57</td>
<td>2,872</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>893</td>
<td>3,262</td>
<td>855</td>
<td>763</td>
<td>-9</td>
<td>54</td>
<td>5,818</td>
</tr>
</tbody>
</table>

### D) Difference between the information for 1980 and 1970

<table>
<thead>
<tr>
<th>State of birth</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>Country</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8,125</td>
<td>-107</td>
<td>39</td>
<td>-39</td>
<td>-22</td>
<td>-5</td>
<td>7,991</td>
</tr>
<tr>
<td>B</td>
<td>67</td>
<td>9,875</td>
<td>218</td>
<td>302</td>
<td>43</td>
<td>2</td>
<td>10,507</td>
</tr>
<tr>
<td>C</td>
<td>216</td>
<td>1,202</td>
<td>10,019</td>
<td>294</td>
<td>41</td>
<td>8</td>
<td>11,780</td>
</tr>
<tr>
<td>D</td>
<td>202</td>
<td>667</td>
<td>-103</td>
<td>2,962</td>
<td>-71</td>
<td>-8</td>
<td>3,649</td>
</tr>
<tr>
<td>E</td>
<td>408</td>
<td>1,500</td>
<td>701</td>
<td>206</td>
<td>20,075</td>
<td>57</td>
<td>22,947</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>9,018</td>
<td>13,137</td>
<td>10,874</td>
<td>3,725</td>
<td>20,066</td>
<td>54</td>
<td>56,874</td>
</tr>
</tbody>
</table>
Procedures based on population characteristics

If the population growth rate in any area of a country is different from the rate of natural increase, the discrepancy is due to migration. Although such an observation is the simplest way to detect migration, the formula can seldom be applied due to the lack of reliable information on natural increase.

Mortality reduces the population of a cohort through time. The size of the reduction depends on the age of the cohort and on the level of mortality in the society. The expected trend in cohort size permits an indirect estimation of the net migration of a population. For instance, suppose a census taken in 1975 enumerated 1,298 persons at ages 45 to 49 in a certain area; and 10 years later, in 1985, a second census enumerated 1,476 persons at ages 55 to 59 in the same area. One may rightly conclude that this cohort was increased by in-migration during the intercensal period. The problem is to know with some accuracy the number of migrants. If the number of deaths pertaining to the cohort were known from vital registration data (perhaps 125 deaths), then the number of migrants would be 1,476 - (1,298 - 125) = 303. Or, following a more logical reasoning, if there were 1,298 persons in 1975 and during the 10 year period 125 persons died, the expected population in 1985 would be 1,298 - 125 = 1,173. But if 1,476 persons were enumerated, the difference between the enumerated and the expected population represents the immigration: 1,476 - 1,173 = 303. Unfortunately, death registration is incomplete in most developing countries, and so an estimated number of migrants based on this technique is generally biased.

When statistics on deaths are lacking, survival ratios from estimated life tables may be used in their place. For instance, if it is known that the cohort of persons 45 to 49 years old enumerated in 1975 had a survival ratio of 0.9037 during the intercensal period, then, multiplying the number of persons in 1975 by the survival ratio yields the expected number of persons in 1985. In the example given above, the expected number is 1,298 x 0.9037 = 1,173, and the difference between the enumerated population in the 1985 census (1,476) and the expected population (1,173) equals the number of migrants (303). The procedure can be applied to each age group, and hence an estimated number of migrants by age can be obtained (see appendix V-1).

The procedure described above not only requires a life table representing the intercensal mortality level, but also assumes that both censuses had the same coverage and that there was no age misreporting. The three requirements are not met in most populations, and hence the estimation of migrants is usually unreliable. If the more recent census is overenumerated or underenumerated to a greater degree than the earlier one, in-migrants will be overestimated or underestimated, respectively. In addition, when applied to states or several areas, this procedure requires life tables for these states or areas, which frequently are not available. Differential census coverage at the state level also produces errors in the estimation of migrants.
Some of the above-mentioned biases resulting from the use of life tables for detecting migration movements can be avoided by applying the same method but using survivorship values from the population census instead of from life tables. As before, this requires the censuses to be 5 or 10 years apart so cohorts can easily be followed during the intercensal period. In addition, the technique is based on several other assumptions: (a) although overall census coverage does not have to be the same at the two dates, the procedure assumes that each area’s coverage was the same; (b) although overall age misreporting does not have to be the same at the two dates, the procedure assumes that each area’s age misreporting was the same; (c) mortality is assumed to be the same in all areas of the country, although this assumption can be avoided if the actual levels of mortality in each area are known.

Some of the survivorship rates calculated based on census information may be greater than 1.0, which would seem to be illogical. This situation could result from different coverage or different "age preference" in the two censuses, or from international migration. Even though unexpected, these values should be used because they take into account such anomalies. Under the assumption that errors are similar in all parts of the country, these values will help to detect the number of internal migrants (if mortality is the same in all areas). When information based on direct questions for detecting migrants is not available, the procedure based on census survivorship is the most highly recommended for estimating internal migration during the intercensal period, particularly for urban and rural areas. (See spreadsheet CSRMIG in appendix V-1, and its documentation in volume II).

The procedure can be improved if information is available on the level of mortality for each area for which migrants are to be estimated. The census survivorship rates for the whole country can be adjusted using the estimated levels in each area. For example, if life tables are available for each area, census survivorship rates can be adjusted by applying the ratios of the survivorship rates pertaining to the life tables of each area to the rates pertaining to the life table for the whole country. Furthermore, if life tables for each area are not available but estimates of life expectancy are, then model life tables can be generated and used in the procedure just described.

The procedure described above for estimating migrants provides information on age but not necessarily age at the moment of migration, as the estimate refers to migrants who are alive at the end of the intercensal period. For instance, persons 55 to 59 years old at the end of the intercensal period could have migrated at any time since the previous census, when they were 45 to 49 years old (assuming a 10-year intercensal period). Correct ages are required for the purpose of calculating migration rates. An acceptable solution to this problem is to assume that the age at the moment of migration was an average of the ages at the beginning and end of the intercensal period. In the example given above, the average age would be 50 to 54 years.

A final problem is that this procedure does not estimate the migration of children born during the intercensal period. Usually this problem is solved based on the assumption that young children migrate with their mothers.
International Migration

The estimation of international migration is more difficult than that of internal migration. It depends not only on information from the country itself but also, in the case of international emigration, on information from other countries as well. Even in countries with excellent statistical systems, information on international migration may be unreliable. An increasing number of international migrants enter a country without proper legal documents, and their entry often goes unrecorded.

Census information on foreign-born population

Censuses collect information on place of birth as well as on nationality. Information on place of birth that includes the foreign born may be used to estimate international migration. In panel C of table V-2, for example, information on the foreign born was included, as provided by responses to a census question on place of birth. These data indicate the intercensal change in the foreign-born population without taking into account intercensal mortality. According to these data, there were 54 more foreign-born persons in 1980 than in 1970. At the state level, 5 and 8 foreigners left states A and D, respectively, and 2, 8, and 57 of them arrived in states B, C and E, respectively. However, it is impossible to know from these data in which states those 54 net migrants (the total of the column) are currently living, as the information reveals only the net movement of the foreign-born population. The data miss not only the total exit and entry of the foreign-born population, but also the international emigration of persons born in the country.

In order to detect the emigration of persons born in the country, information on the foreign-born population in other countries is required. Furthermore, information from all the other countries should pertain to the same year. But not every country of the world provides information on its foreign-born population, and not all countries' censuses are taken the same year, and so international migration is one of the less accurate measures of population change.

If the extent of coverage of population censuses were known, and if a country had a good vital registration system, net international migration could be estimated for both the foreign born and country nationals. The procedure would be the same as the one described for internal migration. That is, based on two censuses and vital statistics, an "expected" population for the latest census date could be estimated and compared with the enumerated population. Any difference would be due to international migration. Although the procedure is logical, however, the technique does not offer good results for the majority of countries due to errors and differential coverage in the censuses. Furthermore, illegal migration adds complications to the process of obtaining accurate measures of international migration.

Some countries collect information on arrivals and departures of passengers at the official borders of the national territory. However, even
for those few countries that process the data, such information is usually unreliable.

In recent years, specific questions for detecting international migration have been tested in the censuses and surveys of a few countries. Questions are posed to designated members of the household, asking whether or not they have a family member living abroad. Techniques based on responses to these questions are in the process of being developed, and there is hope for their improvement (Somoza, 1977; Hill, 1981; Zaba, 1987; Zlotnik, 1988). The techniques are useful in the sense that they detect a certain magnitude of the emigration from a country. However, up to the present, application of these techniques is limited to a very few countries where such questions were included in censuses or surveys. Moreover, false answers and lack of informants' knowledge about their relatives may make the results of these techniques of questionable reliability. Finally, an approach to use the idea of the growth balance equation to detect international migration has recently been developed (Hill, 1989).
Appendix V-1

Estimating Rural-to-Urban Migration During an Intercensal Period

Description

Although the technique can be used to estimate migration from any area, the explanation refers only to rural-to-urban migration. The rural population of one census is projected to the date of a subsequent census. The difference between the projected and enumerated rural populations at the later date serves as an estimate of the number of migrants during the intercensal period.

Data Required

The total and rural population of two consecutive censuses 5 years, 10 years, or 15 years apart, by 5-year age groups.

Assumptions

(1) The degree of completeness of enumeration in each age group is the same in rural areas as in the whole country.

(2) Mortality in rural areas is the same as in the whole country.

Procedure

(1) Intercensal survival ratios are calculated by taking ratios of the national population in each age group as enumerated in the later census to the national population in the same cohort as enumerated in the earlier census.

(2) The rural population of the earlier census is projected using the survival ratios calculated in (1).

(3) The difference between the projected rural population in each age group as derived in (2) and the rural population in the same age group as enumerated in the later census represents the estimated net number of migrants out of (or into) rural areas.

Advantages

(1) The required information is available in most population censuses.

(2) If the actual populations do not depart significantly from the assumptions, the results are not greatly affected by age
misreporting or by differential completeness of enumeration in the two censuses.

(3) The technique provides an approximate age structure of the migrants.

Limitations

(1) If there is a difference in mortality between rural and urban areas, the results will be affected. If mortality is higher in rural than urban areas (and hence higher than in the country as a whole), the estimate of migrants will be: (a) greater than the actual number if it represents out-migration, and (b) smaller than the actual number if it represents in-migration.

(2) Censuses have to be 5 years, 10 years, or 15 years apart. If they are not, the census populations should be shifted so the intercensal periods correspond to one of these alternatives. This change may be a source of error.

(3) Only net migrants are estimated, not the actual flux of in-migrants and out-migrants.

(4) Age of the migrants is not estimated precisely. Assume, for instance, that the intercensal period is 10 years, and we are examining the estimate based on the age group 20 to 24 years at the earlier census. The migrants in this cohort may have migrated anytime between ages 20 years and 34 years. We may assume that they migrated at their mean age during the intercensal period (25 to 29 years), but this is not necessarily correct.

(5) No estimate of migration is obtained for persons born during the intercensal period.

Note

If life tables for the whole country and for rural areas are available, the intercensal survival ratios for the whole population (as mentioned under Procedure, step 1) can be modified. The modification is calculated for each age by taking a ratio of the rural survival ratio to the country survival ratio, and multiplying the result by the country's intercensal survival ratio.

Software

The Bureau of the Census has developed a spreadsheet to apply this technique for intercensal periods of 5 years, 10 years, and 15 years. The spreadsheet is called CSRMIG, and its documentation is presented in volume II.
REFERENCES


Chapter VI
GEOGRAPHIC DISTRIBUTION OF THE POPULATION

This Chapter in Brief

The distribution of the population over the territory of a country can be analyzed using census data. There are several indices that measure particular aspects of population distribution.

(1) A population map presents a visual description of population distribution or population density.

(2) The Gini concentration ratio summarizes the concentration of population over a territory.

(3) The center of population represents the average balance of the population on a territory.

(4) The index of population potential measures the interaction of the population in particular points on the territory.

(5) The index of population dispersion also measures the concentration of population over a territory.

(6) The primacy index measures the concentration of the population of the largest city of the country in relation to that of other cities. See Bureau of the Census spreadsheet URBINDEX in appendix VI-1 and its documentation in volume II.

(7) The city concentration ratio applies the Gini concentration ratio to cities only. See Bureau of the Census spreadsheet URBINDEX in appendix VI-1 and its documentation in volume II.

Spreadsheet and Method That Can Be Used for Analyzing Urbanization and City Distribution, According to the Available Information

<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>City and urban populations</td>
<td>URBINDEX</td>
<td>Urbanization and city indices.</td>
</tr>
</tbody>
</table>

259
Where to Find the Software

The following spreadsheet for measuring the geographic distribution of a population can be found in volume II.

Introduction

The distribution of the population within a country’s territory depends on both historical factors and present-day characteristics. While early inhabitants of the earth tended to locate in a particular territory because of food possibilities, weather conditions, and geological considerations, current populations locate for completely different reasons. The domestication of plants and animals started a process of transforming humans from nomadic groups to sedentary societies. The development of agriculture several millenniums ago permitted the emergence of the first cities because the ability to produce food locally generated the possibility of permanent settlements. Nonetheless, during the greater part of human history, populations were rather broadly dispersed over the inhabitable areas of each continent.

It was only during the last two centuries that the distribution of the population in each country became of concern in the process of planning. During this century, the emerging metropolises and areas of high population density in the countryside of each nation created the need to know and analyze the distribution of the population, as densely settled areas require much more careful planning of public services and more attention to long-term environmental issues.

This chapter discusses ways to describe and measure the distribution of the population, presents the indices most frequently used for such purposes, and considers the analysis and measurement of the process of urbanization.
Population Distribution

Information on place of residence collected in population censuses may be used to describe the distribution of the population over a country's territory. Population data by states or provinces and smaller administrative units within each state or province may be used to describe where the population lives in the country. In addition, population density (number of persons per square kilometer or per square mile) may be used to analyze the concentration of population. Population density is calculated by dividing the total population of an area or political division by the corresponding surface area, usually expressed in square kilometers (see table VI-1).

The analysis of population density requires some knowledge of the topography of the countries or areas being compared, or the results may be misleading. For example, if a large part of a country or region is uninhabited because it is comprised of high mountains, deserts, or lakes, the average density may be quite low, even though most of the people live in crowded conditions in a small part of the region. Hence, it may be more appropriate to estimate population densities omitting surface areas of uninhabitable parts of the territory. For this reason, densities are sometimes calculated as persons per unit of agricultural or arable land. The appropriate measure of density depends on the purpose of the analysis being conducted, but the calculation depends also on the availability of information in the appropriate detail.

Population maps

If the description of the distribution of the population within a country were to be based only on statistical tables and charts, the user or analyst would have to have a perfect knowledge of the administrative divisions of the nation to understand just where the population is located. Although statistical data and indices are essential in the measurement of population distribution, the use of maps is probably the best and easiest way to present the information.

Such maps may reflect either the physical distribution of the nation's people or the relative density of settlement in the various administrative divisions. A population distribution map usually uses one dot to represent each certain number of people (perhaps each 1,000) in both rural and urban areas, with the dots sometimes varying in size to show the relative size of cities. A map showing the population distribution of Senegal and The Gambia is shown in figure VI-1.

A population density map can be made by using an administrative map of the country and information on population density in each administrative area to be considered. To obtain a complete visual perception of the distribution of the population, the map should show rather small areas, at least down to the level of secondary administrative divisions.

To prepare the map, each administrative area is colored or shaded according to its density levels. For a visual effect that makes the map easy
to interpret, the darker the color or the more intense the shaded pattern, the higher the density that it represents (see figure VI-2).
## Table VI-1. Population Densities at the State Level

<table>
<thead>
<tr>
<th>States</th>
<th>Population (1)</th>
<th>Area (2)</th>
<th>Density (3) (1)/(2)</th>
<th>Arable area (4)</th>
<th>Density (5) (1)/(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>8,937,742</td>
<td>75,660</td>
<td>118</td>
<td>53,899</td>
<td>166</td>
</tr>
<tr>
<td>A</td>
<td>3,905</td>
<td>3,254</td>
<td>1</td>
<td>152</td>
<td>26</td>
</tr>
<tr>
<td>B</td>
<td>145,485</td>
<td>5,683</td>
<td>26</td>
<td>3,021</td>
<td>48</td>
</tr>
<tr>
<td>C</td>
<td>155,608</td>
<td>2,871</td>
<td>54</td>
<td>2,654</td>
<td>59</td>
</tr>
<tr>
<td>D</td>
<td>98,512</td>
<td>7,463</td>
<td>13</td>
<td>5,578</td>
<td>18</td>
</tr>
<tr>
<td>E</td>
<td>1,270,147</td>
<td>12,908</td>
<td>98</td>
<td>10,647</td>
<td>119</td>
</tr>
<tr>
<td>F</td>
<td>1,670</td>
<td>2,783</td>
<td>1</td>
<td>32</td>
<td>52</td>
</tr>
<tr>
<td>G</td>
<td>56,192</td>
<td>4,356</td>
<td>13</td>
<td>2,587</td>
<td>22</td>
</tr>
<tr>
<td>H</td>
<td>3,653,643</td>
<td>1,971</td>
<td>1,854</td>
<td>1,971</td>
<td>1,854</td>
</tr>
<tr>
<td>I</td>
<td>157,383</td>
<td>3,245</td>
<td>49</td>
<td>3,098</td>
<td>51</td>
</tr>
<tr>
<td>J</td>
<td>173,837</td>
<td>7,658</td>
<td>23</td>
<td>4,538</td>
<td>38</td>
</tr>
<tr>
<td>K</td>
<td>261,796</td>
<td>4,356</td>
<td>60</td>
<td>4,198</td>
<td>62</td>
</tr>
<tr>
<td>L</td>
<td>68,563</td>
<td>1,265</td>
<td>54</td>
<td>1,176</td>
<td>58</td>
</tr>
<tr>
<td>M</td>
<td>164,410</td>
<td>374</td>
<td>440</td>
<td>354</td>
<td>464</td>
</tr>
<tr>
<td>N</td>
<td>499,043</td>
<td>789</td>
<td>633</td>
<td>701</td>
<td>712</td>
</tr>
<tr>
<td>O</td>
<td>37,572</td>
<td>2,378</td>
<td>16</td>
<td>943</td>
<td>40</td>
</tr>
<tr>
<td>P</td>
<td>1,103,962</td>
<td>5,476</td>
<td>202</td>
<td>5,187</td>
<td>213</td>
</tr>
<tr>
<td>Q</td>
<td>980,261</td>
<td>5,576</td>
<td>176</td>
<td>4,987</td>
<td>197</td>
</tr>
<tr>
<td>R</td>
<td>105,755</td>
<td>3,254</td>
<td>33</td>
<td>2,075</td>
<td>51</td>
</tr>
</tbody>
</table>

*Includes area of cities.

Source: U.S. Bureau of the Census.
Figure VI-1. Population Distribution of Senegal and The Gambia: 1988

Source: U.S. Bureau of the Census.
Figure VI-2. Population Density of Senegal, by Arrondissement: 1988

Source: U.S. Bureau of the Census.
The Gini concentration ratio

In addition to its utility in making maps, information on population density can be used to calculate a summary measure of how evenly or unevenly the population is distributed over the territory of the country. The Gini concentration ratio is such a measure, adapted from other branches of social science (Yntema, 1933). If the population were evenly distributed in a country, a given proportion of the country's territory would have the same proportion of its population; that is, 20 percent of the territory would have 20 percent of the population, and 57 percent of the territory would have 57 percent of the population. Graphically, this situation would be represented by the diagonal of a square, or a 45 degree line (see figure VI-3, line AB). In actual cases, a country's population is never evenly distributed over the territory, and hence the cumulative proportions of area and population will differ one from the other. For example, if the regions of a country are ranked according to their population density (from lowest to highest) and the cumulative percentages of population and area are calculated, each pair of percentages may be as illustrated in figure VI-3. In this figure, the ratio of the shaded portion to the triangle labeled ABC represents the Gini concentration ratio. To calculate this ratio, the shaded portion of the figure must be estimated. The steps are as follows:

![Figure VI-3. Gini Concentration Ratio](image)

Note: The index is the ratio of the shaded area to the area of triangle ABC.

266
(1) Calculate the population density of each administrative division.

(2) Rank the administrative divisions from lowest to highest density.

(3) Cumulate the surface area and the population of the ranked administrative divisions.

(4) Take percentages of the figures cumulated in step (3) in relation to the total surface area and population of the country.

(5) Apply the formula presented in appendix VI-1.

The Gini concentration ratio can be used to analyze the historical population concentration in a country as a whole or the population concentration in each state or province. The higher the value of the index, the higher the concentration of the population within specific areas of the territory.

Although it is a useful measure for certain purposes, this index of population concentration must be interpreted with caution. Usually, the smaller the surface area of the territory being analyzed, the higher the value of the index. If, for example, regions could be defined in such a way that all uninhabited land areas were excluded, then all inhabited land areas would have high population densities, and the index value would be close to its maximum. Thus, the number of subdivisions in a country’s territory has an impact on the results. In a comparative analysis of countries or regions with large differences in the number of administrative divisions, this problem can be overcome by using information for only a selected number of them. For example, information may be selected only for cumulated areas whose percentages in relation to the total country’s area (step (5) in the above procedures) are close to 5, 10, 15, ..., 90, 95, and 100. This process could be considered a "standardization effect" when various countries are being compared.

Other population distribution measures

Some less commonly used indices or concepts related to the distribution or concentration of population are the center of population, the population-potential index, and the index of population dispersion. These concepts are discussed briefly below; for a more detailed description, see Arriaga (1973).

Center of population. This is a summary measure that represents an equilibrium point in the geographic area of the country, derived under the assumptions that the surface area of the country is flat and that each inhabitant weighs the same. Under such conditions, if the country were suspended in space from the population center, the area of the country would be in a horizontal equilibrium. To calculate the center of population, it is necessary to know the geographical coordinates of each inhabited place in the country and the population of each place. For cities, all inhabitants are assumed to have the same geographical coordinates. For dispersed population in an area, it is assumed that all inhabitants have the coordinates of the geographic center of the area. See formula in appendix VI-1.
Population-potential index. This index measures the population interaction in a given territory based on the dual premises that (a) the closer the population is to a particular point, the greater the interaction; and (b) the larger the population, the greater the interaction (Stewart, 1956). For example, the interaction of a city population is higher at the center of the city (the population potential) than at a point 100 kilometers away. The concept may also be applied in a broader sense: assume that two particular points (call them place 1 and place 2) in a country are at equal distances from two cities (call them city A and city B). Place 1 is 80 kilometers from city A, which has 10,000 inhabitants, while place 2 is 80 kilometers from city B, which has 100,000 inhabitants. In this situation, the population potential of place 2 is greater than that of place 1 because, although the distance from its respective city is the same, city B is larger than city A, and a premise states that the larger the population the greater the interaction. Likewise, if a "place 3" were located between two cities with the same population size, its interaction would be greater with the closer of the two cities.

The overall population potential of a specific point X is calculated by dividing the population of all other inhabited places in the country by their distance from point X, and taking the sum of all possible ratios. See formula in appendix VI-1.

The concept of population potential may be used to make maps with iso-population lines; these are lines that connect points with the same population potential in a country.

Index of population dispersion. This index, also known as standard distance, measures the dispersion of the population over a territory (Bachi, 1958). The underlying concept was borrowed from the statistical concept of standard deviation, in this instance represented by the standard distance. While in the statistical concept of standard deviation a mean is required, in the concept of standard distance the "mean" is taken as the center of population. Once the country's center of population is calculated and located on the national territory, the standard distance from it can be calculated as follows:

1. Calculate the square of the distance from the center of population to each inhabited place.
2. Weigh each of the values calculated in (1) by multiplying it by the population of its corresponding inhabited place.
3. Sum the products calculated in (2).
4. Divide the value obtained in (3) by the sum of the total population of all places.
5. Take the square root of the ratio (3)/(4); this is the standard distance, or the index of population dispersion.

The greater the value of the index, the more dispersed the population over the territory of the country. The index of standard distance is 268
sometimes useful in the analysis of historical changes in the population distribution of a country. For international comparisons, however, the total area of each country has an impact on the value of the index; the larger the surface area of the country, the greater the value of the index.

For more details on the indices measuring population distribution, see appendix VI-1.

Location and population of principal cities

Cities are one of the most important aspects in the analysis of the distribution of the population. Once a city achieves a certain population size and economic importance in the country, it increasingly attracts new migrants, and information on cities becomes crucial with respect to policies concerning population distribution. It is necessary to know not only the location and overall population size of each city, but also the age and sex structure of the population. In addition, knowledge of other characteristics, such as labor force, housing, and education, must be integrated into the planning process. The distribution of city population may be analyzed using various indices, of which some of the most commonly used are presented in this section.

Index of city concentration. An urbanization index presented in chapter VII estimates the average city size of the place of residence of the population. The index of city concentration takes the concept of average city size of residence and calculates the percent of the country's population in such a theoretical city (Arriaga, 1973). To calculate the index, first an average is taken of the responses about the size of the city of residence from a random sample of people living in the country. The index represents the proportion of the total population in this average city size. The higher the value of the index, the higher the concentration of the population in large cities. The formula for calculating the index is presented in appendix VI-1, and an example is presented in table A-VI-1.3. The spreadsheet URBINDEX (see volume II) developed at the Bureau of the Census calculates the value of the index.

Index of city distribution. This index was developed as a result of the size-rank rule (Zipf, 1949). This rule was empirically derived from historical observations about the distribution of the population size of cities in developed countries. It was observed that, if a country's cities were ranked by population size, the population of any city (for example, the one ranked k) was approximately equal to the population of the largest city divided by the rank k. In order to better identify relationships among cities, the rank in the formula was modified by an exponent. The result became the now well-known "size-rank rule." (See "Index of city distribution" in appendix VI-1). While in some countries the value of the exponent (z constant) differed significantly from 1, in others its value was close to 1, indicating that, on average, the population of the second largest city was about half that of the largest, the population of the third city was almost one-third that of the largest, and so forth. Observation of this latter relationship led to the creation of primacy indices.
**Primacy index.** The primacy index (Davis, 1962) most frequently used relates the population of the largest city to the combined populations of the second, third and fourth largest cities. The rationale behind this index is the premise that, if cities follow the size-rank rule when the exponent modifying the rank is 1, then the ratio of the largest city to the next three largest cities should be close to 1. This is because the population of the second, third and fourth cities would be one-half, one-third, and one-fourth the population of the largest city. Summing the fractions (one-half plus one-third plus one-fourth) yields a sum slightly greater than the total population of the largest city and, hence, the ratio is actually slightly under 1.

The primacy index is used for comparative analysis. It shows the primacy of the population of the largest city in relation to that of the next three largest cities. In European countries and other developed countries of the world, the ratio tended to fluctuate around unity in the past. However, in developing countries with rather small populations, the ratio fluctuates between 2 and 3, indicating an enormous primacy of the largest city (frequently the capital) over the next most populous cities (see table VI-2).

Other primacy indices can easily be constructed and calculated if cities are grouped in such a way that the total population of each group can be expected to approximate the population of the largest city, under the assumption that the population of the cities follows the size-rank rule without an exponent equal to 1. For instance, under this assumption, the combined population of the following groups of cities should approximate the population of the largest city: second to fourth largest, fifth to twelfth largest, thirteenth to thirty-fourth largest, thirty-fifth to ninety-fourth largest, and so on. (See appendix VI-1 and documentation of spreadsheet URBINDEX in volume II.)

**City concentration ratio.** The Gini concentration ratio was introduced above in the section on analysis of the distribution of the population over the territory. The same index can be used to analyze the distribution of city populations. The application of the Gini concentration ratio to cities (Arriaga, 1977 and 1982) resulted in a simplified formula. The interpretation of the index is the same as when it was applied to geographic areas: if all cities had the same population, then, for a total of 100 cities, the population of 15 cities would have 15 percent of the total city population; the population of 39 cities would have 39 percent of the total city population, and so forth. As in the case of geographic area analysis, this situation would be represented by the diagonal of a square, or a 45 degree line as represented by line AB in figure VI-3.
Table VI-2. Primacy Index for Selected Countries and Years

<table>
<thead>
<tr>
<th>Country and year</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria, 1977</td>
<td>1.49</td>
</tr>
<tr>
<td>Argentina, 1980</td>
<td>3.92</td>
</tr>
<tr>
<td>Chile, 1985</td>
<td>3.97</td>
</tr>
<tr>
<td>China, 1982</td>
<td>.46</td>
</tr>
<tr>
<td>Cuba, 1985</td>
<td>2.53</td>
</tr>
<tr>
<td>Denmark, 1985</td>
<td>3.06</td>
</tr>
<tr>
<td>France, 1982</td>
<td>2.67</td>
</tr>
<tr>
<td>India, 1981</td>
<td>.50</td>
</tr>
<tr>
<td>Uganda, 1985</td>
<td>1.66</td>
</tr>
<tr>
<td>United States, 1984</td>
<td>.68</td>
</tr>
</tbody>
</table>

Example of how the index is calculated:

For France, the population of the four largest cities was:

<table>
<thead>
<tr>
<th>City</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris</td>
<td>8,510,000</td>
</tr>
<tr>
<td>Lyon</td>
<td>1,170,000</td>
</tr>
<tr>
<td>Marseille</td>
<td>1,080,000</td>
</tr>
<tr>
<td>Lille</td>
<td>935,000</td>
</tr>
</tbody>
</table>

The index is: $\frac{8,510,000}{(1,170,000+1,080,000+935,000)} = 2.67$

Source: Calculated from U.N. Demographic Yearbook 1986
See appendix VI-1.

Actual cities, however, have different populations: for example, 10 percent of the cities (from the smallest to the largest) may have 1 percent of the total city population; 20 percent of the smallest cities, 2.5 percent of the population; 30 percent of the smallest cities, 8 percent of the population, and so on. Hence, a graph representing the city concentration ratio may resemble figure VI-3. As in the case of the Gini concentration ratio, the ratio of the shaded portion to the area of the triangle labeled ABC is the city concentration ratio.

As in the case of geographic area analysis, the value of this index is affected by the number of cities included. To avoid such effects in comparative analyses, two possibilities are suggested; each has a different meaning but may help in the analysis of city distributions. One possibility is to select the same number of cities, for example, the 10 largest cities, in each of the populations to be analyzed. In this case, the index would reflect the concentration of city population among the 10 largest cities in each country or region.

The other possibility is to calculate the index for all cities over a selected size in each of the populations to be analyzed, for example, all cities with a population of 100,000 or more inhabitants. In this case, the index would reflect the concentration of city population among the cities of 100,000 or more inhabitants in each country or population (see table VI-3 and
appendix VI-1). A spreadsheet, URBINDEX, is available for making the calculation of this index (see documentation in volume II).

Several other indices have been developed for measuring the concentration of population in cities. They are presented in various publications (Arriaga, 1973), and they do not differ appreciably from the indices presented here.

Table VI-3. Gini Concentration Ratio for Cities of 100,000 Inhabitants or More: Selected Countries and Years

<table>
<thead>
<tr>
<th>Country and year</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria, 1977</td>
<td>.529</td>
</tr>
<tr>
<td>Argentina, 1980</td>
<td>.658</td>
</tr>
<tr>
<td>Chile, 1985</td>
<td>.637</td>
</tr>
<tr>
<td>Cuba, 1985</td>
<td>.556</td>
</tr>
<tr>
<td>Denmark, 1985</td>
<td>.526</td>
</tr>
<tr>
<td>France, 1982</td>
<td>.579</td>
</tr>
<tr>
<td>Uganda, 1985</td>
<td>.410</td>
</tr>
</tbody>
</table>

Note: See calculation example in Appendix table A-VI-1.3.

Source: Calculated from U.N. Demographic Yearbook 1986.
Appendix VI-1

Measuring Population Distribution

Indices Frequently Used

There are several indices to measure population distribution and dispersion over a territory. The more frequently used indices are presented here.

Gini concentration ratio (CR)

To calculate this ratio (Yntema, 1933), population density within each subdivision of a country is ranked from lowest to highest. Then the corresponding populations and areas of each subdivision are cumulated. Finally, the percentages of these cumulations are taken, and the following formula is applied:

\[ CR = 10^{-4} \left( \sum_{k=2}^{m} PP_k \cdot PA_{k-1} - \sum_{k=2}^{m} PP_{k-1} \cdot PA_k \right) \]

Where:

- \( PP \) and \( PA \) represent the percentages of population and area cumulated in each subdivision of the country;
- \( k \) refers to the rank pertaining to each of these subdivisions; and
- \( m \) is the total number of subdivisions.

The higher the ratio, the more concentrated the population of the country. This index is affected by the size of the areas used in the calculation. The smaller the area, the higher the value of the index. For historical comparisons, the same areas should be maintained. See calculation in table A-VI-1.1.

Center of population (PC)

The center of population in a territory is the geographical point where the center of gravity of the population is located. It is obtained by calculating geographical coordinates whose longitude is a weighted average of the geographical longitudes of all inhabited places, and whose latitude is a weighted average of the latitudes of all inhabited places. In symbols:

\[ PC (LA, LO) \]
### Table A-VI-1.1. Calculation of the Gini Concentration Ratio

<table>
<thead>
<tr>
<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>F</td>
<td>1,670</td>
<td>2,783</td>
<td>.6</td>
<td>1</td>
<td>1,670</td>
<td>2,783</td>
<td>.02</td>
<td>3.68</td>
<td>.2</td>
</tr>
<tr>
<td>A</td>
<td>3,905</td>
<td>3,254</td>
<td>1.2</td>
<td>2</td>
<td>5,575</td>
<td>6,037</td>
<td>.06</td>
<td>7.98</td>
<td>5.5</td>
</tr>
<tr>
<td>G</td>
<td>56,192</td>
<td>4,356</td>
<td>12.9</td>
<td>3</td>
<td>61,767</td>
<td>10,393</td>
<td>.69</td>
<td>13.74</td>
<td>24.6</td>
</tr>
<tr>
<td>D</td>
<td>98,512</td>
<td>7,463</td>
<td>13.2</td>
<td>4</td>
<td>160,279</td>
<td>17,856</td>
<td>1.79</td>
<td>25.60</td>
<td>52.2</td>
</tr>
<tr>
<td>O</td>
<td>37,572</td>
<td>2,378</td>
<td>15.8</td>
<td>5</td>
<td>197,851</td>
<td>20,234</td>
<td>2.21</td>
<td>26.74</td>
<td>81.6</td>
</tr>
<tr>
<td>J</td>
<td>173,837</td>
<td>7,658</td>
<td>22.7</td>
<td>6</td>
<td>371,688</td>
<td>27,892</td>
<td>4.16</td>
<td>36.86</td>
<td>213.3</td>
</tr>
<tr>
<td>B</td>
<td>145,485</td>
<td>5,683</td>
<td>25.6</td>
<td>7</td>
<td>517,172</td>
<td>33,575</td>
<td>5.79</td>
<td>44.38</td>
<td>281.7</td>
</tr>
<tr>
<td>R</td>
<td>105,755</td>
<td>3,254</td>
<td>32.5</td>
<td>8</td>
<td>622,927</td>
<td>36,829</td>
<td>6.97</td>
<td>48.68</td>
<td>369.2</td>
</tr>
<tr>
<td>L</td>
<td>157,383</td>
<td>3,245</td>
<td>48.5</td>
<td>9</td>
<td>780,310</td>
<td>40,074</td>
<td>8.73</td>
<td>52.97</td>
<td>549.6</td>
</tr>
<tr>
<td>C</td>
<td>155,608</td>
<td>2,871</td>
<td>54.2</td>
<td>10</td>
<td>935,918</td>
<td>42,945</td>
<td>10.47</td>
<td>56.76</td>
<td>611.9</td>
</tr>
<tr>
<td>K</td>
<td>68,563</td>
<td>1,265</td>
<td>54.2</td>
<td>11</td>
<td>1,004,481</td>
<td>44,210</td>
<td>11.24</td>
<td>58.43</td>
<td>721.4</td>
</tr>
<tr>
<td>E</td>
<td>261,796</td>
<td>4,356</td>
<td>60.1</td>
<td>12</td>
<td>1,266,277</td>
<td>48,566</td>
<td>14.17</td>
<td>64.19</td>
<td>1151.1</td>
</tr>
<tr>
<td>Q</td>
<td>1,270,147</td>
<td>12,908</td>
<td>98.4</td>
<td>13</td>
<td>2,536,424</td>
<td>61,474</td>
<td>28.38</td>
<td>81.25</td>
<td>3196.9</td>
</tr>
<tr>
<td>P</td>
<td>980,261</td>
<td>5,576</td>
<td>175.8</td>
<td>14</td>
<td>3,516,685</td>
<td>67,050</td>
<td>39.35</td>
<td>88.62</td>
<td>5481.5</td>
</tr>
<tr>
<td>M</td>
<td>1,103,962</td>
<td>5,476</td>
<td>201.6</td>
<td>15</td>
<td>4,620,646</td>
<td>72,526</td>
<td>51.70</td>
<td>95.86</td>
<td>5132.0</td>
</tr>
<tr>
<td>N</td>
<td>164,410</td>
<td>374</td>
<td>439.6</td>
<td>16</td>
<td>4,785,057</td>
<td>72,900</td>
<td>53.54</td>
<td>96.35</td>
<td>5696.5</td>
</tr>
<tr>
<td>H</td>
<td>499,043</td>
<td>789</td>
<td>632.5</td>
<td>17</td>
<td>5,284,099</td>
<td>73,689</td>
<td>59.12</td>
<td>97.39</td>
<td>7939.5</td>
</tr>
<tr>
<td></td>
<td>3,653,643</td>
<td>1,971</td>
<td>1853.7</td>
<td>18</td>
<td>8,937,742</td>
<td>75,660</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**Totals**: 8,937,742 75,660 118.1

Calculation of the Gini Concentration Ratio:

\[
CR = \frac{33329.79 - 26356.4 \times 0.001}{33329.79} = 0.6973
\]

**Source**: Table VI-1.
Where:

LA and LO represent the average latitudes and longitudes.
These values are calculated as:

\[ LA = \frac{\sum P_i \times LA_i}{P} \]
\[ LO = \frac{\sum P_i \times LO_i}{P} \]

Where:

LA\(_i\) and LO\(_i\) represent the latitude and longitude of place \(i\);
P\(_i\) represents the population of place \(i\); and
P represents the total population of the territory.

**Population-potential (PP) Index**

The population-potential index (Stewart, 1956) for a particular place \(a\) is calculated as follows:

\[ PP_a = \sum \left( \frac{P_i}{D_i} \right) \]

Where:

\(P_i\) is the population of place \(i\); and
\(D_i\) is the distance from place \(a\) to place \(i\).

The sum of all possible ratios (from any place in relation to place \(a\)) gives the population potential for place \(a\). Repeating the procedure for other places, population potential can be calculated for most places in a country. Iso-population lines connect places with the same population potential. Places with high levels of population potential are areas affected by large populations.

**Index of population dispersion (PD)**

This index (Bachi, 1958) is calculated as a standard deviation of the distances from the center of population to any other inhabited place in a territory. In symbols:

\[ PD = \frac{\sum P_i D_i^2}{P} \]

Where:

\(P\) and \(P_i\) are the total population and the population of place \(i\); and
$D_i$ is the distance from the center of population to place $i$.

This distance is calculated as:

$$D_i = \sqrt{K^2 (LO - LO_i)^2 + Q^2 (LA - LA_i)^2}$$

Where:

- $LA$ and $LA_i$ represent the latitude of the center of population and of place $i$, respectively;
- $LO$ and $LO_i$ represent the longitude of the center of population and of place $i$, respectively; and
- $K$ and $Q$ are constants used to convert geographical degrees into distance units.

The lower the value of this index, the more concentrated the population of a country.

**Index of city distribution $(C)$**

This index (Zipf, 1949) is derived from the size-rank rule. This rule is:

$$C_k = C_1 x. k^{-z}$$

Where:

- $C_1$ is the largest city;
- $C_k$ is the $k$th largest city;
- $k$ is the rank of the city according to its size; and
- $z$ is a constant that results from fitting the size-rank rule to the distribution of cities.

The value of $z$ is called the index of city distribution; this index varies from country to country. The higher the value of the index the more concentrated the city population in the largest cities. The index $z$ is calculated using the least square technique as follows:

$$z = \frac{\sum \ln (k) \cdot \ln (C_1 / C_k)}{\sum \ln (k)^2}$$

Where the symbols are the same as above. The spreadsheet URBINDEX of the Bureau of the Census calculates this index. See its documentation in volume II.
**Primacy index (PI)**

The primacy index (Davis, 1962) is calculated as follows:

\[ \text{PI}_4 = \frac{c_1}{(c_2 + c_3 + c_4)} \]

Where:

- \( c \) represents the city population; and
- subscripts 1, 2, 3, and 4 represent the first, second, third, and fourth largest cities, respectively.

An alternative is to take into account the 11th largest city, as follows:

\[ \text{PI}_{11} = \frac{2 \times c_1}{(c_2 + c_3 + \ldots + c_{11})} \]

Where the symbols are the same as above but involve more cities, from the 1st to the 11th largest, as indicated by the subscripts.

For instance, in the examples given in table A-VI-1.2, the index for Algeria was calculated as:

\[ \text{PI}_4 = \frac{1,740,461}{(543,485 + 378,668 + 246,049)} \]
\[ = 1.490 \]

The higher the value of this index, the more concentrated the population in the largest city. See examples in table A-VI-1.2. The spreadsheet URBINDEX of the Bureau of the Census calculates this index. See its documentation in volume II.

**Index of city concentration (CC)**

This index (Arriaga, 1973) is the percent of the total population that theoretically is represented by the average size of place of residence of a country's population. (See index of mean city population size in chapter VII.) The higher the value of the index, the higher the concentration of the population in large cities. The index is calculated as:

\[ \text{CC} = 100 \times \sum \frac{c_i^2}{p^2} \]

Where:

- \( c_i^2 \) is the square of the population size of city \( i \); and
- \( p^2 \) is the square of the total population of the country.
Table A-VI-1.2. Calculation of the Primacy Index for Selected Countries

<table>
<thead>
<tr>
<th>Country, Year</th>
<th>City 1</th>
<th>Population 1</th>
<th>City 2</th>
<th>Population 2</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria, 1977</td>
<td>Alg</td>
<td>1,740,461</td>
<td>Shanghai</td>
<td>11,185,100</td>
<td>1.48986305</td>
</tr>
<tr>
<td></td>
<td>Oram</td>
<td>543,485</td>
<td>Beijing</td>
<td>9,179,660</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constantine</td>
<td>378,668</td>
<td>Tianjin</td>
<td>7,790,160</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annaba</td>
<td>266,049</td>
<td>Changing</td>
<td>6,511,130</td>
<td></td>
</tr>
<tr>
<td>Tanzania 1985</td>
<td>Dar es Salaam</td>
<td>1,096,000</td>
<td>Calcutta</td>
<td>9,194,018</td>
<td>1.66060606</td>
</tr>
<tr>
<td></td>
<td>Mbawa</td>
<td>252,000</td>
<td>Bombay</td>
<td>8,243,405</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tabora</td>
<td>214,000</td>
<td>Delhi</td>
<td>5,729,283</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mbeya</td>
<td>194,000</td>
<td>Madras</td>
<td>4,289,347</td>
<td></td>
</tr>
<tr>
<td>Cuba, 1985</td>
<td>La Habana</td>
<td>2,014,806</td>
<td>Kobenhavn</td>
<td>1,358,540</td>
<td>2.52537652</td>
</tr>
<tr>
<td></td>
<td>Santiago de Cuba</td>
<td>358,764</td>
<td>Arhus</td>
<td>194,348</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Camaguey</td>
<td>260,782</td>
<td>Odense</td>
<td>136,803</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Santa Clara</td>
<td>178,278</td>
<td>Alborg</td>
<td>113,865</td>
<td></td>
</tr>
<tr>
<td>Argentina, 1980</td>
<td>Buenos Aires</td>
<td>10,728,000</td>
<td>Paris</td>
<td>8,510,000</td>
<td>3.91675794</td>
</tr>
<tr>
<td></td>
<td>Córdoba</td>
<td>1,055,000</td>
<td>Lyon</td>
<td>1,170,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rosario</td>
<td>1,016,000</td>
<td>Marseille</td>
<td>1,080,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mendoza</td>
<td>668,000</td>
<td>Lille</td>
<td>935,000</td>
<td></td>
</tr>
<tr>
<td>Chile, 1985</td>
<td>Santiago</td>
<td>4,099,714</td>
<td>New York</td>
<td>17,807,100</td>
<td>3.96988295</td>
</tr>
<tr>
<td></td>
<td>Valparaíso-Viña</td>
<td>534,331</td>
<td>Los Angeles</td>
<td>12,372,600</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Concepción</td>
<td>280,713</td>
<td>Chicago</td>
<td>8,035,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Talcahuano</td>
<td>217,660</td>
<td>Philadelphia</td>
<td>5,755,300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Index</td>
<td>Index</td>
<td>Index</td>
<td>Index</td>
<td></td>
</tr>
</tbody>
</table>

For practical purposes, it is not necessary to take into account every place of residence in a country. Cities over a selected size (for example, 10,000 inhabitants and over, 25,000 inhabitants and over, 50,000 inhabitants and over, and so on) may be used. See example in table A-VI-1.3. The spreadsheet URBINDEX of the Bureau of the Census calculates this index. See documentation in volume II.

**City concentration ratio (CCR)**

This index (Arriaga, 1977) was developed based on the previous index for areas. However, this one uses cities instead of areas, and the calculations do not require percentages or the cumulation of city populations. The formula is:

\[
CCR = \frac{(n-1)}{n} - \frac{2 \times \sum (k-1) \cdot C_k}{n \cdot CP}
\]

Where:

- \(n\) is the number of cities;
- \(k\) is the rank of the city (1st is largest; nth, smallest);
- \(C_k\) is the population of the city whose rank is \(k\); and
- \(CP\) is the total population of all cities considered.

The higher the value of the ratio, the higher the concentration of population in the largest cities. See example in table A-VI-1.4. The spreadsheet URBINDEX of the Bureau of the Census calculates this index. See documentation in volume II.

**Software**

Some of these indices are included in a spreadsheet, URBINDEX, developed by the Bureau of the Census. See documentation in volume II.
Table A-VI-1.3. Calculation of the Index of City Concentration: Chile, 1985

<table>
<thead>
<tr>
<th>Cities</th>
<th>Population (thousands)</th>
<th>Square C(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santiago</td>
<td>4,100</td>
<td>16,810,000</td>
</tr>
<tr>
<td>Valparaíso-Viña</td>
<td>534</td>
<td>205,156</td>
</tr>
<tr>
<td>Concepción</td>
<td>281</td>
<td>78,961</td>
</tr>
<tr>
<td>Talcahuano</td>
<td>218</td>
<td>47,324</td>
</tr>
<tr>
<td>Antofagasta</td>
<td>203</td>
<td>41,209</td>
</tr>
<tr>
<td>Temuco</td>
<td>168</td>
<td>28,224</td>
</tr>
<tr>
<td>Arica</td>
<td>158</td>
<td>24,964</td>
</tr>
<tr>
<td>Rancagua</td>
<td>157</td>
<td>24,649</td>
</tr>
<tr>
<td>Talca</td>
<td>138</td>
<td>19,044</td>
</tr>
<tr>
<td>San Bernardo</td>
<td>136</td>
<td>18,496</td>
</tr>
<tr>
<td>Iquique</td>
<td>127</td>
<td>16,129</td>
</tr>
<tr>
<td>Chillán</td>
<td>127</td>
<td>16,129</td>
</tr>
<tr>
<td>Puente Alto</td>
<td>125</td>
<td>15,625</td>
</tr>
<tr>
<td>Punta Arenas</td>
<td>107</td>
<td>11,449</td>
</tr>
<tr>
<td>Valdivia</td>
<td>105</td>
<td>11,025</td>
</tr>
<tr>
<td>Osorno</td>
<td>102</td>
<td>10,404</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>6,786</strong></td>
<td><strong>17,458,988</strong></td>
</tr>
</tbody>
</table>

Total population of the country: 12,122,146,942,884

Index of City Concentration = \( \frac{100 \times 17,458,988}{146,942,884} = 11.88 \)

Table A-VI-1.4. Calculation of the City Concentration Ratio Index, Chile, 1982

<table>
<thead>
<tr>
<th>Cities</th>
<th>Rank</th>
<th>Population C(k)</th>
<th>Products k.C(k+1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santiago</td>
<td>1</td>
<td>4,099,714</td>
<td></td>
</tr>
<tr>
<td>Valparaiso-Viña</td>
<td>2</td>
<td>534,331</td>
<td>534,331</td>
</tr>
<tr>
<td>Concepción</td>
<td>3</td>
<td>280,713</td>
<td>561,426</td>
</tr>
<tr>
<td>Talcahuano</td>
<td>4</td>
<td>217,660</td>
<td>652,980</td>
</tr>
<tr>
<td>Antofagasta</td>
<td>5</td>
<td>203,067</td>
<td>812,268</td>
</tr>
<tr>
<td>Temuco</td>
<td>6</td>
<td>168,120</td>
<td>860,600</td>
</tr>
<tr>
<td>Arica</td>
<td>7</td>
<td>158,422</td>
<td>950,532</td>
</tr>
<tr>
<td>Rancagua</td>
<td>8</td>
<td>157,209</td>
<td>1,100,463</td>
</tr>
<tr>
<td>Talca</td>
<td>9</td>
<td>137,621</td>
<td>1,100,968</td>
</tr>
<tr>
<td>San Barnardo</td>
<td>10</td>
<td>136,224</td>
<td>1,226,016</td>
</tr>
<tr>
<td>Iquique</td>
<td>11</td>
<td>127,491</td>
<td>1,274,910</td>
</tr>
<tr>
<td>Chillán</td>
<td>12</td>
<td>126,581</td>
<td>1,392,391</td>
</tr>
<tr>
<td>Puente Alto</td>
<td>13</td>
<td>125,297</td>
<td>1,503,564</td>
</tr>
<tr>
<td>Punta Arenas</td>
<td>14</td>
<td>107,064</td>
<td>1,391,832</td>
</tr>
<tr>
<td>Valdivia</td>
<td>15</td>
<td>104,910</td>
<td>1,668,740</td>
</tr>
<tr>
<td>Osorno</td>
<td>16</td>
<td>101,948</td>
<td>1,529,220</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>6,786,372</td>
<td>16,340,241</td>
</tr>
</tbody>
</table>

City Concentration Index = \((15/16) \cdot (2 \times 16,340,241) / (16 \times 6,786,372)\) = 0.6365

REFERENCES


Chapter VII

URBANIZATION

This Chapter in Brief

Urbanization occurs as a result of people's choosing to live concentrated in cities. There are two aspects of the urbanization process: the increasing proportion of people selecting urban areas of residence, and the growth of cities.

Degree of urbanization

Several indices measure particular aspects of the degree of urbanization at a given moment. Some of them are:

1. Percent of population in urban areas.
2. The urban/rural ratio.
3. Mean city population size.

Tempo of urbanization

The tempo of urbanization measures the change in the level of urbanization by analyzing changes in the indices used for measuring the degree of urbanization.

Where to Find the Software

The following spreadsheet for measuring urbanization can be found in volume II.

Bureau of the Census spreadsheet:

URBINDEX
Introduction

Urbanization is a twentieth century demographic phenomenon. More and more people are changing residence from rural to urban areas, and an increasing proportion of them are selecting large cities. These events produce two aspects of urbanization whose measurement should be differentiated: (a) an increase in the percent of population living in urban areas; and (b) growth of the city population.

Measuring Urbanization

Although it is generally understood that urbanization is the concentration of population in cities, there is no consensus among scholars regarding some of the basic issues. These include the very definition of an urban locality and how to achieve the geographic delimitation of a city. The definitions of urban locality vary from country to country. Some are based only on the size of the population residing in the locality, while others specify the presence of some characteristics that authorities in the country consider proper to an urban area, such as paved roads, high schools, electricity, or self-administration. The varying urban definitions among countries of the world create problems in international comparisons but, in actual practice, concepts and size of urban populations do not vary significantly according to the various definitions.

For a particular country, once the characteristics of an urban locality have been defined, there still exists the problem of delimiting the physical boundaries of each urban place. To establish where a city ends is not an easy task. In most cases, a solution to the problem is to consider population densities of the areas surrounding a city, or sometimes the average distance between dwellings.

Definitions aside, demographers and planners are concerned with how to measure the urbanization process. For this, it is necessary to have base data and some means to measure change. Fortunately, there seems to be general agreement that the urbanization process concerns the number of persons living in urban localities in relation to the country's total population. Thus, the simplest index to measure the urbanization process in a population is the percent of the total population living in urban localities.

Percent urban

This index refers to the number of persons living in urban localities for each 100 people living in the country. It is calculated by taking the ratio of the urban population to the total population of the country, times 100.

The easy interpretation of this index is its great advantage. However, in comparative analyses, it is questionable whether it reflects the relative levels of urbanization among countries. For example, in 1975, 83 percent of the population of Uruguay but only 75 percent of the population of Japan lived in urban localities. An analysis of urban characteristics of the two
countries almost certainly would show that, in most aspects of the urbanization process, Japan was "more urban" than Uruguay.

A further disadvantage of the percent urban as a measure is that once a country or an area of a country achieves a high proportion of urban population, further increases in the percent urban are negligible, although the "urbanization process" may continue in the sense that the size of cities continues to increase (see table VII-1).

Urban/rural ratio

Another index proposed for measuring the urbanization process is the ratio of the urban to the rural population. This index is also easy to understand: it gives the number of urban dwellers for each person living in rural areas. Thus, if the index has a value of 1, it means that the urban population is equal to the rural population. The advantage of this index is that, unlike the percent urban, the urban/rural ratio does not have an upper limit. It may vary from zero to virtually the total population of the country assuming, for calculation purposes, that there is at least one person living in rural areas.

The urban/rural ratio is useful for estimating changes in the urbanization process. In a comparative analysis of countries or of areas within a country, the proportion urban and the urban/rural ratio are related. Although the calculated values of the two indices would be different in magnitude, the ranking of the degree of urbanization of the countries or areas would be the same (see table VII-2).

Mean city population size

Both of the indices described above are easy to understand, but neither one takes into account an important dimension of the urbanization process: the size of cities. For instance, in the above example of Japan and Uruguay, the population of Tokyo is several times the population of Montevideo. Furthermore, several other cities in Japan are larger than the largest city in Uruguay. Although the two countries may have a similar percent of their populations in urban areas, their urbanization cannot be as alike as the percent urban would suggest. Since urban problems are related to the size of urban localities, for specific analyses it would be desirable to take into account the size of cities in establishing an urbanization index.

For this purpose, an index called mean city population size was created (Arriaga, 1973). It takes into account the size of cities and consequently gives different results when applied to countries or areas where the percent urban is the same but the size of cities is different. The value of this index represents the average city size of residence of the population. In other words, it tells the size of the city in which the average person lives.
<table>
<thead>
<tr>
<th>Year</th>
<th>Percent Urban</th>
<th>Average Annual Change</th>
<th>Year</th>
<th>Percent Urban</th>
<th>Average Annual Change</th>
<th>Year</th>
<th>Percent Urban</th>
<th>Average Annual Change</th>
<th>Year</th>
<th>Percent Urban</th>
<th>Average Annual Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td>28.3</td>
<td>.17</td>
<td>1911</td>
<td>45.4</td>
<td>.41</td>
<td>1901</td>
<td>43.1</td>
<td>.49</td>
<td>1790</td>
<td>5.1</td>
<td>.09</td>
</tr>
<tr>
<td>1930</td>
<td>33.5</td>
<td>.16</td>
<td>1917</td>
<td>49.5</td>
<td>.22</td>
<td>1911</td>
<td>47.9</td>
<td>.59</td>
<td>1800</td>
<td>6.1</td>
<td>.12</td>
</tr>
<tr>
<td>1960</td>
<td>42.6</td>
<td>.82</td>
<td>1926</td>
<td>53.7</td>
<td>.06</td>
<td>1921</td>
<td>53.9</td>
<td>.51</td>
<td>1810</td>
<td>7.3</td>
<td>.01</td>
</tr>
<tr>
<td>1950</td>
<td>50.7</td>
<td>.83</td>
<td>1931</td>
<td>54.3</td>
<td>.86</td>
<td>1926</td>
<td>56.4</td>
<td>.04</td>
<td>1820</td>
<td>7.2</td>
<td>.16</td>
</tr>
<tr>
<td>1970</td>
<td>58.7</td>
<td>.73</td>
<td>1941</td>
<td>62.9</td>
<td>.74</td>
<td>1936</td>
<td>56.8</td>
<td>.36</td>
<td>1830</td>
<td>8.8</td>
<td>.20</td>
</tr>
<tr>
<td>1980</td>
<td>66.3</td>
<td></td>
<td>1951</td>
<td>66.6</td>
<td>.75</td>
<td>1945</td>
<td>60.3</td>
<td>-.94</td>
<td>1840</td>
<td>10.8</td>
<td>.45</td>
</tr>
<tr>
<td>1990</td>
<td>75.5</td>
<td>-.11</td>
<td>1956</td>
<td>55.2</td>
<td>.03</td>
<td>1950</td>
<td>55.1</td>
<td>1.71</td>
<td>1850</td>
<td>15.3</td>
<td>.45</td>
</tr>
<tr>
<td>1976</td>
<td>75.5</td>
<td>.04</td>
<td>1961</td>
<td>63.6</td>
<td>.94</td>
<td>1960</td>
<td>66.3</td>
<td>2.64</td>
<td>1860</td>
<td>19.8</td>
<td>.59</td>
</tr>
<tr>
<td>1981</td>
<td>75.7</td>
<td></td>
<td>1971</td>
<td>81.4</td>
<td>.32</td>
<td>1970</td>
<td>68.3</td>
<td>2.86</td>
<td>1870</td>
<td>25.7</td>
<td>.25</td>
</tr>
<tr>
<td>1981</td>
<td>83.6</td>
<td></td>
<td>1976</td>
<td>85.0</td>
<td>.12</td>
<td>1980</td>
<td>73.7</td>
<td></td>
<td>1880</td>
<td>28.2</td>
<td>.69</td>
</tr>
</tbody>
</table>

Source: Calculated from population censuses.
Table VII-2. Urban/Rural Ratio and Difference Between Urban and Rural Annual Growth Rates: Selected Countries

<table>
<thead>
<tr>
<th>Year</th>
<th>Urban/Rural Ratio</th>
<th>Rural Annual Growth Rate</th>
<th>Urban Annual Growth Rate</th>
<th>Urban/Rural Difference in Annual Growth Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>Canada</td>
<td>New Zealand</td>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>.39</td>
<td>.54</td>
<td>1.03</td>
<td>1.42</td>
</tr>
<tr>
<td>1930</td>
<td>.50</td>
<td>.74</td>
<td>3.4</td>
<td>3.1</td>
</tr>
<tr>
<td>1940</td>
<td>.50</td>
<td>.98</td>
<td>3.6</td>
<td>3.2</td>
</tr>
<tr>
<td>1950</td>
<td>.74</td>
<td>1.16</td>
<td>.3</td>
<td>2.1</td>
</tr>
<tr>
<td>1960</td>
<td>1.03</td>
<td>1.19</td>
<td>3.6</td>
<td>3.2</td>
</tr>
<tr>
<td>1970</td>
<td>1.42</td>
<td>1.70</td>
<td>3.2</td>
<td>3.1</td>
</tr>
<tr>
<td>1980</td>
<td>1.96</td>
<td>2.00</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>1990</td>
<td>.54</td>
<td>1.17</td>
<td>2.1</td>
<td>1.05</td>
</tr>
<tr>
<td>2000</td>
<td>.76</td>
<td>1.32</td>
<td>2.1</td>
<td>1.23</td>
</tr>
<tr>
<td>2010</td>
<td>2.0</td>
<td>2.7</td>
<td>.2</td>
<td>2.16</td>
</tr>
<tr>
<td>2020</td>
<td>2.4</td>
<td>3.2</td>
<td>3.2</td>
<td>2.16</td>
</tr>
<tr>
<td>2030</td>
<td>2.1</td>
<td>3.1</td>
<td>3.1</td>
<td>2.16</td>
</tr>
</tbody>
</table>

Source: Calculated from population censuses.
The concept should not be confused with the average size of cities. For example, suppose a country has just two cities, with populations of 10,000 and 100,000, respectively. The average size of the two cities is 55,000. However, most of the people live in the city of 100,000, and hence, the average city size of residence of the 110,000 inhabitants is not 55,000 but 91,818 inhabitants.

An analogy with the mean age of a group of persons may clarify the concept. Suppose there are 10 persons of age 20 years and 100 persons of age 40 years. The average age of the whole group of 110 persons is not 30 years (because there are more persons in the older age group), but the result of a weighted average, or 38.2 years.

The mean city population size index has the same underlying concept as the mean age example, and its calculation is similar. In the case of the mean age, the population at each age is multiplied by the age; in the case of the mean city population size, the population of each city is multiplied by the size of the city. In this case, the size of the city is also the population of the city, so the calculation is easy: the population of each city is multiplied by itself, the products are added, and the sum is divided by the total population of the country. If all inhabited places of the country (including isolated persons) are taken into account, statistically this index is the expected value of the city size of residence of the population.

As described above, the calculation of the index would require information on every inhabited place in the country. However, a summary version takes into account only cities over a certain convenient size. The value of the index is not significantly affected by the choice of a particular minimum population size of cities to be included in the calculation. This is an advantage because varying definitions of the urban concept thus do not significantly affect the value of the index. The index also can be calculated from aggregated data of cities by city-size categories (Arriaga, 1973).

The summarized version of the index can be interpreted as a product of two aspects of urbanization: the average size of residence of the population that is considered urban (living in cities above the cut-off size used in the calculation) times the proportion of population in such cities or urban areas (see formula in appendix VII-1). The advantage of considering these two factors as part of the index is that it allows the researcher to analyze whether or not the change in urbanization is due mainly to population increase in existing cities or to the emergence of cities newly classified as urban during the time period of analysis (see appendix VII-1).

This index may be useful for comparing the urbanization of states within a country and for analyzing some social aspects of urbanization that are strongly related to city size or to the urbanization of particular social groups. It may also be used for international comparisons of city size. In cases where states or countries have the same proportion of population living in cities over a certain size, this index detects differences in the size of cities. For example, in 1980, Mexico and Belgium had practically the same proportion of population living in cities of 25,000 or more inhabitants. However, not only did Mexico have several cities larger than Brussels, but
Mexico City was almost nine times larger than the capital of the European country. Therefore, although the proportion of population in cities of 25,000 or more inhabitants was the same, the size of the cities was different. This difference in size of cities is measured by the index of mean city population size. While the value of the index for Mexico was 3,076 thousands, for Belgium it was 345 thousands. See appendix VII-1 and documentation of the spreadsheet URBINDEX in volume II.

These results do not mean that Mexico is almost nine times more urbanized than Brussels. They mean that the population of Mexico, on average, resides in cities of larger size than the population of Belgium does.

**Tempo of Urbanization**

The previous section presented some of the indices for measuring the level of urbanization in a population. In this section, another dimension of the urbanization process is presented: the measurement of how rapidly urbanization takes place, or what is known as "urbanization tempo."

The measurement of urbanization tempo indicates the pace at which a specific area is urbanizing. If the degree or level of urbanization in a country or area is known for two or more dates, the tempo is measured by the annual change in the index used for calculating the level of urbanization. Although potentially useful, this procedure of measuring tempo may require some caution depending on the index used for calculating the level of urbanization. If the level were calculated using the percent urban, for example, areas achieving a high degree of urbanization (close to 100 percent) would show a slowing down of the tempo, not necessarily because the urbanization process was slowing down but because the percent urban has the limits 0 to 100. Hence, it is not recommended to measure the tempo of urbanization by the annual change in the percent urban, (see table VII-1).

One of the best indices for measuring the tempo of urbanization is the difference between the annual population growth rates of urban and rural areas. For instance, if the annual population growth rates during an intercensal period for urban and rural populations were 4.6 percent and 1.5 percent, respectively, the urbanization tempo would be 3.1.

The urban/rural ratio is related to the difference between the two mentioned rates. If the urban/rural ratio is known for more than one date, the annual exponential growth rate of the urban/rural ratio is also the difference between the urban and rural annual population growth rates.

Measuring the tempo of urbanization not only makes sense for determining how fast one area is growing in relation to the other but also is related to the indices for measuring the level of urbanization. Both the percent urban and the urban/rural ratios are related to the difference between the urban and rural population growth rates. For example, it has been observed empirically that the percent urban follows a trend resembling an S-shaped curve. At low levels of percent urban, the increase is slow; at levels of about 50 percent urban, it is frequently the fastest; and at higher levels of percent urban,
the change slows down. Analytically, an S-shaped curve can be represented by a logistic function. If a logistic function is fitted to the changes in percent urban, with lower and upper asymptotes of 0 and 100 respectively, the growth rate of the logistic is the same as the difference between the urban and rural population growth rates (Arriaga, 1973). The higher the logistic rate, the shorter the time needed for achieving high levels of percent urban (see appendix VII-2).

For the reasons mentioned above, it is advisable to measure the tempo of urbanization as the difference between the annual growth rates of the urban and rural populations.

City Population Growth

An analysis of the urbanization process should include a study of the urban population growth by size of city. Data on individual cities or aggregations of cities of similar size can be used in the analysis. The growth of cities can be studied from two points of view: (a) by considering the same cities during an intercensal period; or (b) by considering all cities within the same classification by size of city. For example, the study can observe the cities of a certain size at an earlier (or later) census date and follow the population change of the same cities to a later (or earlier) census date. The second possibility is to aggregate the information on all cities in the same city-size categories in both censuses (not necessarily the same cities) and analyze the change of the population in the categories. Although both alternatives should be analyzed, the first one is more frequently used. Developing countries face a rapid city population growth; its analysis requires information on the principal cities in each province or state.

The urbanization process can be characterized by the growth of cities, that is, whether the largest cities are growing faster than small cities, or vice versa. As mentioned in an earlier section, the mean city population size index permits an analysis of such situations during a specified period. Such an index (see details in appendix VII-1) can be decomposed into two components: the proportion of the population in urban localities, and the mean city population size of the urban population (not of the whole country). If the change in the index during an intercensal period is due mainly to the size of urban localities, it probably would imply that large cities are growing faster than small ones, or that no significant number of new cities is entering into the classification of urban localities. On the other hand, if most of the change in the index is due to the percent urban, it means that the increase in the percent urban has been due mainly to reclassification of localities from rural to urban, or to the faster growth of small cities than large ones (Arriaga, 1973) (see appendix VII-1).

Components of Urban Growth

Urbanization seems to occur when the population residing in localities designated as urban is growing faster than that in the rest of the country. But the questions are: Why does the urban population grow faster than the
rural? In which area are crude birth and death rates higher or lower? How significant is the component of rural-to-urban net migration? Is the urban population growing mainly because of the existing urban localities or because a significant number of rural localities have been redefined as urban? Answers to these questions require an analysis of the components of urban growth. The analysis should take into account two points of view: from the point of view of the city itself, urban increase occurs by growth of existing urban localities, by reclassification of cities (from rural to urban or vice versa), and by annexation of new territory to existing cities. From the point of view of population growth, or an analysis of the components of change, urban population grows because of natural increase, because of migration (rural to urban or vice versa) and because some people are reclassified (that is, their residence becomes urban by annexation when the place where they live is reclassified as urban). The three alternatives from each point of view generate nine combinations of factors, of which only eight are feasible (see Table VII-3). The information needed to fill in all possibilities in the table is difficult to obtain because it requires not only census data but also vital statistics registration. However, marginals of the table can be estimated based on census data.

### Table VII-3. Components of Urban Growth

<table>
<thead>
<tr>
<th>Components of change</th>
<th>Type of locality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same urban localities</td>
</tr>
<tr>
<td>Natural change</td>
<td></td>
</tr>
<tr>
<td>Migration</td>
<td></td>
</tr>
<tr>
<td>Reclassified population</td>
<td>Not suitable</td>
</tr>
</tbody>
</table>

Appendix VII-1

Index of Mean City Population Size

As discussed in the text, this index (Arriaga, 1973) is based on the same concept as the mean age of the population, but applied in this case to the city size of residence of the population. The formula is:

\[ MC = \frac{\sum_{i=1}^{n} C_i^2}{n} \frac{1}{P} \]  

(A)

Where:

- \( C_i \) is the population of city \( i \);
- \( n \) is the total number of cities; and
- \( P \) is the total population of the country.

It has been proved that if, instead of all cities, only those over a certain size are considered, the index is affected only slightly (Arriaga, 1973). Hence, taking only those cities over a certain size (such as 20,000 and over or 50,000 and over), the index can be decomposed into two factors:

\[ MC = \frac{\sum_{i=1}^{u} C_i^2}{u} \frac{1}{\sum_{i=1}^{u} C_i} \frac{1}{P} \]  

(B)

Where \( u \) is the number of cities over any chosen limit.

An example of how this index is calculated is presented in table A-VII-1.1. The two factors can be used for analyzing the change in the degree of urbanization over time.

If the city data are grouped into city size categories, the index can be calculated as follows:

(1) If the number of cities in each category is known:

\[ MC_a = \frac{\sum_{j=1}^{s} \frac{K_j^2}{m_j}}{\sum_{j=1}^{s} \frac{1}{m_j}} \frac{1}{P} \]  

(C)
Where:

\[ K_j \] is the population in the city-size category \( j \);

\[ m_j \] is the number of cities in the city-size category \( j \);

\( s \) is the number of city-size categories; and

\( P \) is the total population.

(2) If the number of cities in each category is not known:

\[
MC_n = \frac{\sum_{j=1}^{s} K_j z_j}{P} \tag{D}
\]

Where:

\[ z_j \] is the geometric mean of the limits of the city-size category; and

other symbols are the same as in the previous formula.

If the largest city-size category is open-ended and the number of cities included in it is unknown, an estimate of the average city size should be used.

This index incorporates the size of city into the measurement of urbanization. In comparing two countries where the proportion of the population in cities over a certain size is similar but the city sizes are different, this index will detect the discrepancy. For instance, in the case of Belgium and Mexico: both countries have the same proportion of population in cities of 50,000 inhabitants and over. However, not only is Mexico City almost nine times larger than Brussels, but other cities in Mexico are much larger than those in Belgium (see table A-VII-1.1). The measurement index does not mean that Mexico is almost nine times more urbanized than Belgium. Rather, the interpretation should be that, although both countries have the same proportion of population in cities of 50,000 inhabitants and over, Mexicans live in much larger cities.

Another use of the index of mean city population size is to explain a change in the degree of urbanization. Sometimes the increase of the urban population during a period of time is concentrated in the large cities, while at other times it is concentrated in the small cities. The change in this index can be decomposed into the two components.
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Country’s total</td>
<td>9,849</td>
</tr>
<tr>
<td>Population of cities</td>
<td></td>
</tr>
<tr>
<td>Bruxelles</td>
<td>1,404</td>
</tr>
<tr>
<td>Antwerpen</td>
<td>860</td>
</tr>
<tr>
<td>Charleroi</td>
<td>524</td>
</tr>
<tr>
<td>Liège</td>
<td>462</td>
</tr>
<tr>
<td>Gent</td>
<td>282</td>
</tr>
<tr>
<td>Mons</td>
<td>170</td>
</tr>
<tr>
<td>Kortrijk</td>
<td>137</td>
</tr>
<tr>
<td>Brugge</td>
<td>134</td>
</tr>
<tr>
<td>Aalst</td>
<td>88</td>
</tr>
<tr>
<td>Leuven</td>
<td>88</td>
</tr>
<tr>
<td>Roeselare</td>
<td>83</td>
</tr>
<tr>
<td>Mechelen</td>
<td>72</td>
</tr>
<tr>
<td>Ostende</td>
<td>71</td>
</tr>
<tr>
<td>Namur</td>
<td>59</td>
</tr>
<tr>
<td>Genk</td>
<td>56</td>
</tr>
<tr>
<td>Verviers</td>
<td>56</td>
</tr>
<tr>
<td>Turnhout</td>
<td>52</td>
</tr>
<tr>
<td>Dendermonde</td>
<td>51</td>
</tr>
<tr>
<td>Hasselt</td>
<td>50</td>
</tr>
<tr>
<td>City total</td>
<td>4,699</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mexico 1980 Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country’s total</td>
</tr>
<tr>
<td>Population of cities</td>
</tr>
<tr>
<td>Ciudad De Mexico</td>
</tr>
<tr>
<td>Guadalajara</td>
</tr>
<tr>
<td>Monterrey</td>
</tr>
<tr>
<td>Puebla</td>
</tr>
<tr>
<td>Léon</td>
</tr>
<tr>
<td>Ciudad Juarez</td>
</tr>
<tr>
<td>Torreón</td>
</tr>
<tr>
<td>Tijuana</td>
</tr>
<tr>
<td>San Luis Potosí</td>
</tr>
<tr>
<td>Tampico</td>
</tr>
<tr>
<td>Mérida</td>
</tr>
<tr>
<td>Chihuahua</td>
</tr>
<tr>
<td>Mexicali</td>
</tr>
<tr>
<td>Toluca</td>
</tr>
<tr>
<td>Culiacán</td>
</tr>
<tr>
<td>Veracruz</td>
</tr>
<tr>
<td>Acapulco</td>
</tr>
<tr>
<td>Morelia</td>
</tr>
<tr>
<td>Hermosillo</td>
</tr>
<tr>
<td>Aguascalientes</td>
</tr>
<tr>
<td>Saltillo</td>
</tr>
<tr>
<td>Cuernavaca</td>
</tr>
<tr>
<td>Durango</td>
</tr>
<tr>
<td>Querétaro</td>
</tr>
<tr>
<td>Jalapa</td>
</tr>
<tr>
<td>Nuevo Laredo</td>
</tr>
<tr>
<td>Mazatlán</td>
</tr>
<tr>
<td>Orizaba</td>
</tr>
<tr>
<td>Reinosá</td>
</tr>
<tr>
<td>Matamoros Tamaulipas</td>
</tr>
<tr>
<td>Poza Rica</td>
</tr>
<tr>
<td>Irapuato</td>
</tr>
<tr>
<td>Oaxaca</td>
</tr>
<tr>
<td>Ciudad Obregón</td>
</tr>
<tr>
<td>Villahermosa</td>
</tr>
<tr>
<td>Monclova</td>
</tr>
<tr>
<td>Tepic</td>
</tr>
<tr>
<td>Coatzacoalcos</td>
</tr>
<tr>
<td>Celaya</td>
</tr>
<tr>
<td>Ciudad Victoria</td>
</tr>
<tr>
<td>Tuxtla</td>
</tr>
<tr>
<td>Campeche</td>
</tr>
<tr>
<td>Uruapan</td>
</tr>
<tr>
<td>Ensenada</td>
</tr>
<tr>
<td>Los Mochis</td>
</tr>
</tbody>
</table>
Table A-VII-1.1. Index of Mean City Population Size: Belgium 1981 and Mexico 1980 -- Continued
(Population in thousands)

<table>
<thead>
<tr>
<th>City of Mexico</th>
<th>Population of cities:</th>
<th>Index:</th>
<th>Proportion of population in cities:</th>
<th>Mean urban city population size in thousands:</th>
<th>Index of mean city population size in thousands:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zamora</td>
<td>117</td>
<td></td>
<td>.477</td>
<td>722</td>
<td>345</td>
</tr>
<tr>
<td>Córdoba</td>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pachuaca</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minatitlán</td>
<td>107</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zacatecas</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colima</td>
<td>103</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salamanca</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La Paz</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tapachula</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tehuacán</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piedras Negras</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Luis Rio Colorado</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texmelucán</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hidalgo del Parral</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carmen</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ciudad Mante</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iguale</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chilpancingo</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valles</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nogales</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delicias</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Navojoa</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texpan</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ciudad Guzmán</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moroleón</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chetumal</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zacatelco</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresnillo</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apatzingán</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rio Bravo</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guaymas</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tulancingo</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlixco</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La Piedad Cavadas</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuautla</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City total</td>
<td>31,964</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Belgium: .477  Mexico: .478
Belgium: 722  Mexico: 6,430
Belgium: 345  Mexico: 3,076
Taking logarithms of the index in formula (B) of this appendix, the product of the two components is transformed into two addends. The value of the index is calculated for times 1 and 2, and then the difference between the values is taken:

\[
\ln M_{C_2} = \ln \left[ \frac{\sum_{i=1}^{u} C_{2,i}}{u} \right] + \ln \left[ \frac{\sum_{i=1}^{u} C_{2,i}}{P_2} \right]
\]

minus

\[
\ln M_{C_1} = \ln \left[ \frac{\sum_{i=1}^{u} C_{1,i}}{u} \right] + \ln \left[ \frac{\sum_{i=1}^{u} C_{1,i}}{P_1} \right]
\]

\[
A = B + C
\]

Where:

- \( A \) represents the total change of the index;
- \( B \) represents the part of the index change due to the increase in the size of cities; and
- \( C \) represents the part of the index change due to the increase of the urban population.

\( B \) and \( C \) can be expressed as a percent of the total change \( A \).

Examples of different patterns of urbanization are presented below based on information from Peru (table A-VII-1.2) and Kenya (table A-VII-1.3). In Peru, the number of cities of 50,000 inhabitants and over increased from 14 to 22 during the intercensal period 1972 to 1981, while in Kenya the corresponding increase was from 2 to 5 cities from 1969 to 1979. The question about the importance of each of the two components of the urbanization process is answered by decomposing the index of city population size.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lima-Callao</td>
<td>1,783.7</td>
<td>3,254.8</td>
<td>4,563.7</td>
<td>3,181,586</td>
<td>10,593,723</td>
<td>20,827,358</td>
</tr>
<tr>
<td>Arequipa</td>
<td>166.9</td>
<td>509.1</td>
<td>652.9</td>
<td>27,856</td>
<td>95,543</td>
<td>205,118</td>
</tr>
<tr>
<td>Trujillo</td>
<td>103.0</td>
<td>240.3</td>
<td>354.6</td>
<td>10,609</td>
<td>57,744</td>
<td>125,741</td>
</tr>
<tr>
<td>Chiclayo</td>
<td>95.7</td>
<td>187.8</td>
<td>280.2</td>
<td>9,158</td>
<td>35,269</td>
<td>78,512</td>
</tr>
<tr>
<td>Chimbote</td>
<td>60.0</td>
<td>160.4</td>
<td>216.4</td>
<td>3,600</td>
<td>25,728</td>
<td>44,929</td>
</tr>
<tr>
<td>Piura</td>
<td>72.1</td>
<td>126.0</td>
<td>186.4</td>
<td>5,198</td>
<td>15,876</td>
<td>34,745</td>
</tr>
<tr>
<td>Iquitos</td>
<td>57.8</td>
<td>110.2</td>
<td>173.6</td>
<td>3,341</td>
<td>12,144</td>
<td>30,137</td>
</tr>
<tr>
<td>Cuzco</td>
<td>79.9</td>
<td>121.5</td>
<td>170.9</td>
<td>6,384</td>
<td>14,762</td>
<td>29,207</td>
</tr>
<tr>
<td>Huancayo</td>
<td>64.2</td>
<td>126.8</td>
<td>165.1</td>
<td>4,122</td>
<td>16,078</td>
<td>27,258</td>
</tr>
<tr>
<td>Ica</td>
<td>259.9</td>
<td>111.1</td>
<td>167.1</td>
<td>67,548</td>
<td>12,343</td>
<td>26,563</td>
</tr>
<tr>
<td>Tacna</td>
<td>56.5</td>
<td>92.6</td>
<td>132</td>
<td>3,102</td>
<td>8,575</td>
<td>12,343</td>
</tr>
<tr>
<td>Pucallpa</td>
<td>57.1</td>
<td>92.0</td>
<td>126</td>
<td>3,260</td>
<td>8,446</td>
<td>12,343</td>
</tr>
<tr>
<td>Puno</td>
<td>86.4</td>
<td>80.9</td>
<td>120</td>
<td>3,588</td>
<td>6,545</td>
<td>12,343</td>
</tr>
<tr>
<td>Sullana</td>
<td>59.9</td>
<td>78.0</td>
<td>104</td>
<td>6,084</td>
<td>10,036</td>
<td>14,762</td>
</tr>
<tr>
<td>Juliaca</td>
<td>68.5</td>
<td>68.5</td>
<td>104</td>
<td>4,692</td>
<td>4,692</td>
<td>13,968</td>
</tr>
<tr>
<td>Ayacucho</td>
<td>52.6</td>
<td>62.9</td>
<td>127</td>
<td>2,767</td>
<td>3,956</td>
<td>13,968</td>
</tr>
<tr>
<td>Huacho</td>
<td>60.2</td>
<td>60.2</td>
<td>120</td>
<td>3,624</td>
<td>3,624</td>
<td>10,872</td>
</tr>
<tr>
<td>Cajamarca</td>
<td>59.0</td>
<td>55.1</td>
<td>53.4</td>
<td>3,481</td>
<td>3,036</td>
<td>2,852</td>
</tr>
<tr>
<td>Pisco</td>
<td>52.6</td>
<td>52.6</td>
<td>52.6</td>
<td>2,767</td>
<td>2,767</td>
<td>2,767</td>
</tr>
<tr>
<td>Huanuco</td>
<td>907</td>
<td>13,538</td>
<td>17,031</td>
<td>10,947,223</td>
<td>21,478,789</td>
<td></td>
</tr>
</tbody>
</table>

Total population of the country, in thousands:

<table>
<thead>
<tr>
<th>Year</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1961</td>
<td>9,907</td>
</tr>
<tr>
<td>1972</td>
<td>13,538</td>
</tr>
<tr>
<td>1981</td>
<td>17,031</td>
</tr>
</tbody>
</table>

Source: Calculated from population censuses.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nairobi</td>
<td>266.8</td>
<td>509.3</td>
<td>827.8</td>
<td>71,182</td>
<td>259,386</td>
<td>685,253</td>
</tr>
<tr>
<td>Mombasa</td>
<td>179.6</td>
<td>247.1</td>
<td>341.1</td>
<td>32,256</td>
<td>61,058</td>
<td>116,349</td>
</tr>
<tr>
<td>Kisumu</td>
<td>152.6</td>
<td></td>
<td></td>
<td>23,287</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakuru</td>
<td>92.9</td>
<td></td>
<td></td>
<td>8,630</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eldoret</td>
<td>50.5</td>
<td></td>
<td></td>
<td>2,550</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>446.4</td>
<td>756.4</td>
<td>1464.9</td>
<td>103,438</td>
<td>320,445</td>
<td>836,069</td>
</tr>
</tbody>
</table>

Total population of the country, in thousands:

<table>
<thead>
<tr>
<th>Year</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1962</td>
<td>8,636</td>
</tr>
<tr>
<td>1969</td>
<td>10,943</td>
</tr>
<tr>
<td>1979</td>
<td>15,327</td>
</tr>
</tbody>
</table>

Source: Calculated from population censuses.
The index and its two components are calculated as follows, taking as an example data for Kenya in 1979 from table A-VII-1.3:

(1) Proportion of population in cities:
\[
\frac{1,464.9}{15,327} = 0.0956
\]

(2) Mean urban population size:
\[
\frac{836,069}{1,464.9} = 570.7
\]

(3) Index of city population size:
\[
0.0956 \times 570.7 = \frac{836,069}{15,327} = 54.5
\]

The change in the index of mean city population size is the result of both the increasing size of the cities and the increasing proportion of the population living in cities. Taking logarithms of the index and its two components as described above, the situation of Peru (using data in table A-VII-1.2) is as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>Index (thousands)</th>
<th>Proportion of population in cities</th>
<th>Urban city population size (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>808.6</td>
<td>.3784</td>
<td>2,136.9</td>
</tr>
<tr>
<td>1981</td>
<td>1,262.2</td>
<td>.4413</td>
<td>2,857.6</td>
</tr>
</tbody>
</table>

Logarithms of the above values:

<table>
<thead>
<tr>
<th>Year</th>
<th>Index</th>
<th>Proportion of population in cities</th>
<th>Urban city population size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>6.6953</td>
<td>.9718</td>
<td>7.6671</td>
</tr>
<tr>
<td>1981</td>
<td>7.1398</td>
<td>.8179</td>
<td>7.9577</td>
</tr>
</tbody>
</table>

Logarithms of values for 1972 subtracted from those for 1981:

<table>
<thead>
<tr>
<th>Percent</th>
<th>.4445</th>
<th>.1539</th>
<th>.2906</th>
</tr>
</thead>
<tbody>
<tr>
<td>100.0</td>
<td>34.6</td>
<td>65.4</td>
<td></td>
</tr>
</tbody>
</table>

The interpretation of these results is as follows: during the intercensal period in Perú, the increase in the size of cities existing in 1972 contributed more to the urbanization process (as measured by this index) than did the increase in the proportion of the country's population living in cities (65 percent compared with 35 percent, respectively).

Similar calculations for Kenya yield different results:
<table>
<thead>
<tr>
<th>Year</th>
<th>Index (thousands)</th>
<th>Proportion of population in cities</th>
<th>Urban city population size (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1969</td>
<td>29.3</td>
<td>0.0691</td>
<td>423.6</td>
</tr>
<tr>
<td>1979</td>
<td>54.5</td>
<td>0.0956</td>
<td>570.7</td>
</tr>
</tbody>
</table>

Logarithms of the above values:

<table>
<thead>
<tr>
<th>Year</th>
<th>Value</th>
<th>Proportion of population in cities</th>
<th>Urban city population size (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1969</td>
<td>3.3770</td>
<td>-2.6719</td>
<td>6.0489</td>
</tr>
<tr>
<td>1979</td>
<td>3.9991</td>
<td>-2.3478</td>
<td>6.3469</td>
</tr>
</tbody>
</table>

Logarithms of values for 1969 subtracted from those for 1979:

<table>
<thead>
<tr>
<th>Percent</th>
<th>.6221</th>
<th>.3241</th>
<th>.2980</th>
</tr>
</thead>
</table>

For Kenya, the increase in the proportion of population living in cities was more important than the increase of the size of cities in relation to the urbanization process during the intercensal period 1969 to 1979.
Appendix VII-2

Relationship Between the Logistic Growth Rate and the Growth of Urban and Rural Population

A modified logistic function (Arriaga, 1973) for any index has the format of:

$$I_n = L + \frac{V}{1 + \frac{U-I_0}{I_0-L} e^{wn}}$$

Where:

- $L$ and $U$ are the lower and upper asymptotes;
- $V$ equals $U - L$;
- $I_0$ is an observed value of an index at time 0;
- $I_n$ is an estimated value of the index, following a logistic shape from time 0 to time n;
- Subscripts 0 and n represent different dates;
- $w$ is the growth rate of the logistic; and
- $e$ is the base of natural logarithms.

The growth rate of this modified logistic function can be calculated if, knowing the two asymptotic values, there are two observations of the index: $I_s$ and $I_0$ for time $-s$ and 0. The formula for calculating the growth rate is:

$$w = -\frac{1}{s} \ln \left( \frac{U-I_0}{U-I_s} / \frac{I_0-L}{I_s-L} \right)$$

Where $s$ is the period of time between the two observations.

The formulas presented here allow the fitting of a logistic to any pair of values of an index, provided that the asymptotes are known. In the specific case of fitting a logistic to the percent urban with asymptotes of 0 and 100 (minimum and maximum values of the percentage), the growth rate of the logistic function is also the difference between the urban and rural population growth rates with opposite signs.
For instance, the following information on urban and rural population illustrates the above statement.

<table>
<thead>
<tr>
<th>Population</th>
<th>1974</th>
<th>1987</th>
<th>Annual growth rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>4,458,500</td>
<td>6,063,400</td>
<td>0.0396552</td>
</tr>
<tr>
<td>Urban</td>
<td>1,747,732</td>
<td>2,926,587</td>
<td>0.0105095</td>
</tr>
<tr>
<td>Rural</td>
<td>2,710,768</td>
<td>3,107,613</td>
<td>0.0291458</td>
</tr>
<tr>
<td>Percent urban</td>
<td>39.2</td>
<td>48.5</td>
<td></td>
</tr>
<tr>
<td>Difference between urban and rural growth rates</td>
<td></td>
<td></td>
<td>0.0291458</td>
</tr>
</tbody>
</table>

The calculation of the growth rate of the logistic is performed by applying the above formula, where:

\[ U = 100; L = 0; I_s = 39.2; I_0 = 48.5; \text{ and } s = 13. \]

Based on these values,

\[
\begin{align*}
1 & \quad w = \frac{1}{13} \ln \left( \frac{1.0618557}{1.5510204} \right) \\
\text{ } & \quad - \left( \frac{\ln 0.68461747}{13} \right) = -0.0291458
\end{align*}
\]

For applications of the logistic function to life expectancies at birth, total fertility rates, or any other index, see chapter VIII, appendix VIII-4.
Appendix VII-3

Urban, Rural, and City Populations With Selected Urban Indices

Table VII-3. Urban, Rural, and City Populations with Selected Urbanization Indices

<table>
<thead>
<tr>
<th>Population</th>
<th>1970</th>
<th>1980</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>20,347,581</td>
<td>22,187,564</td>
</tr>
<tr>
<td>Urban</td>
<td>18,134,983</td>
<td>19,936,341</td>
</tr>
<tr>
<td>Rural</td>
<td>2,212,598</td>
<td>2,251,223</td>
</tr>
<tr>
<td>Urban cities:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100,000 and over</td>
<td>14,678,693</td>
<td>16,240,400</td>
</tr>
<tr>
<td>Other</td>
<td>3,456,290</td>
<td>3,695,941</td>
</tr>
<tr>
<td>Cities 100,000 and over:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>5,234,687</td>
<td>5,976,438</td>
</tr>
<tr>
<td>B</td>
<td>2,098,034</td>
<td>2,664,578</td>
</tr>
<tr>
<td>C</td>
<td>1,657,356</td>
<td>1,975,230</td>
</tr>
<tr>
<td>D</td>
<td>853,654</td>
<td>862,129</td>
</tr>
<tr>
<td>E</td>
<td>803,456</td>
<td>859,076</td>
</tr>
<tr>
<td>F</td>
<td>697,345</td>
<td>734,549</td>
</tr>
<tr>
<td>G</td>
<td>576,489</td>
<td>602,388</td>
</tr>
<tr>
<td>H</td>
<td>423,765</td>
<td>422,645</td>
</tr>
<tr>
<td>I</td>
<td>289,456</td>
<td>295,834</td>
</tr>
<tr>
<td>J</td>
<td>256,123</td>
<td>264,734</td>
</tr>
<tr>
<td>K</td>
<td>227,456</td>
<td>202,858</td>
</tr>
<tr>
<td>L</td>
<td>210,568</td>
<td>202,434</td>
</tr>
<tr>
<td>M</td>
<td>198,624</td>
<td>201,765</td>
</tr>
<tr>
<td>N</td>
<td>187,625</td>
<td>194,567</td>
</tr>
<tr>
<td>O</td>
<td>176,254</td>
<td>175,934</td>
</tr>
<tr>
<td>P</td>
<td>165,234</td>
<td>171,956</td>
</tr>
<tr>
<td>Q</td>
<td>156,321</td>
<td>155,634</td>
</tr>
<tr>
<td>R</td>
<td>125,456</td>
<td>128,345</td>
</tr>
<tr>
<td>S</td>
<td>116,548</td>
<td>115,987</td>
</tr>
<tr>
<td>T</td>
<td>113,678</td>
<td>115,834</td>
</tr>
<tr>
<td>U</td>
<td>110,564</td>
<td>117,465</td>
</tr>
<tr>
<td>Urbanization indices:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent urban</td>
<td>89.1</td>
<td>89.8</td>
</tr>
<tr>
<td>Urban/rural ratio</td>
<td>8.20</td>
<td>8.86</td>
</tr>
<tr>
<td>Mean city population size (thousands)</td>
<td>1,837</td>
<td>2,196</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of the Census.
Chapter VIII

POPULATION PROJECTIONS

This Chapter in Brief

Population projection methodology can be divided into two main categories: (a) procedures for projecting the population considering mortality, fertility, and migration, by age and sex (component method); and (b) procedures for projecting the population using mathematical functions applied to population figures but not to each of the components.

Component method

A base population is determined that agrees with known demographic characteristics of the country. Levels of mortality, fertility, and migration are determined for the base year and projected to future years. Then the base population is projected into the future according to the projected components of change.

(1) A base population is obtained after analyzing the available age and sex structure (see chapter II). Bureau of the Census spreadsheets GRPOP-YB, PYRAMID, SINGAGE, AGESEX, AGESMTH, BASEPOP, and BPSTRNG. Those underlined are strongly recommended.

(2) Two aspects of mortality have to be projected: the level and the pattern by age and sex. Bureau of the Census spreadsheets EOLGST, FITLGSTC, LOGISTIC, INTPLTM, and INTPLTF. Also the United Nations MATCH program in the MORTPAK package. Those underlined are strongly recommended.

(3) Two aspects of fertility have to be projected: the level and the age pattern. Bureau of the Census spreadsheets TFRILGST, TFRSINE, FITLGSTC, LOGISTIC, and ASFRPATT. Those underlined are strongly recommended.

(4) If more than one area within a country is to be projected, two aspects of internal migration have to be estimated for the past and projected into the future: the total number of migrants and their age and sex composition.

(5) Two aspects of international migration have to be estimated: the total number of migrants and their age and sex composition.

With the projected components of population growth, the population can be projected by age, sex, and areas if desired. The calculations are performed by a special program developed by the Bureau of the Census, called RUP.
Mathematical functions

Populations, particularly those of small areas without adequate information, can be projected using mathematical functions and procedures. Sometimes population totals are projected, while other times, if control totals exist, proportions of population in relation to the totals are projected. The most common mathematical functions for such projections are linear, exponential and logistic functions.

Spreadsheets and Methods That Can Be Used for Projecting Populations and Components, According to the Available Information

<table>
<thead>
<tr>
<th>Information</th>
<th>Spreadsheet</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age-specific data on population, mortality and fertility.</td>
<td>BASEPOP</td>
<td>Smooths base population consistent with mortality and fertility.</td>
</tr>
<tr>
<td></td>
<td>BPSTRNG</td>
<td>Same as BASEPOP but uses a stronger smoothing process.</td>
</tr>
<tr>
<td>Life expectancies at birth</td>
<td>EOLGST</td>
<td>Interpolates and extrapolates life expectancies at birth.</td>
</tr>
<tr>
<td>Total fertility rates</td>
<td>TFRLGST</td>
<td>Interpolates and extrapolates total fertility rates (logistic).</td>
</tr>
<tr>
<td></td>
<td>TFRSINE</td>
<td>Interpolates and extrapolates total fertility rates (sine).</td>
</tr>
<tr>
<td></td>
<td>ASFRPATT</td>
<td>Estimates age-specific fertility rates.</td>
</tr>
<tr>
<td>Life table</td>
<td>INTPLTM</td>
<td>Interpolates a male life table.</td>
</tr>
<tr>
<td></td>
<td>INTPLTF</td>
<td>Interpolates a female life table.</td>
</tr>
<tr>
<td>Other indices or information</td>
<td>LOGISTIC</td>
<td>Interpolates and extrapolateslogistically.</td>
</tr>
<tr>
<td></td>
<td>FITILGSTC</td>
<td>Interpolates and extrapolateslogistically.</td>
</tr>
<tr>
<td></td>
<td>CTBL32</td>
<td>Adjusts subpopulations to totals.</td>
</tr>
</tbody>
</table>

Note: A special computer program, called RUP, performs the population projection.
Where to Find the Software

The following spreadsheets for use in making population projections can be found in volume II.

Bureau of the Census spreadsheets:

AGESEX   EOLGST   LOGISTIC
AGESMTH   FITLGSTC  PYRAMID
ASFRPATT  GRPOP-YB  SINGAGE
BASEPOP   INTPLTF   TFRLGST
BPSTRNG   INTPLTM   TFRSINE
CTRL32

Note: A special computer program, called RUP, performs the population projection.

Introduction

Previous chapters described and analyzed the basic demographic variables producing change in a population. For a country as a whole, these variables are fertility, mortality, and international migration. For regions or areas within a country, internal migration also contributes to the change in population size and its age and sex structure.

The purpose of a population projection is to anticipate the changes in population size and characteristics. The size of a population can be projected by taking into account changes that have occurred in the past. Population growth rates are usually calculated based on past information, and they are sometimes used in specific mathematical functions for projecting the probable future size of the population. This procedure provides only the total population size, not its characteristics.

A more useful procedure for projecting a population is to simulate how the population changes according to its components of growth: mortality, fertility, and migration. Based on past information, assumptions are made about future trends in these components of change. Then, the projected rates are applied to the age and sex structure of the population, in a simulation taking into account that people die according to their sex and age, that women have children, and that some people change their residence. This procedure for projecting a population is called the component method.
This chapter presents both types of population projections: first, the component method, including some procedures for projecting mortality, fertility, and migration; and then some methods for projecting the total population based on mathematical formulas.

**Component Method**

The component method of projecting a population follows each cohort of people of the same age throughout its lifetime according to its exposure to mortality, fertility, and migration. Starting with a base population by sex and age, the population at each specific age is exposed to the chances of dying as determined by projected mortality levels and patterns by sex and age. Once the deaths are estimated, they are subtracted from the surviving population, and those remaining alive become older. Fertility rates are projected and applied to the female population in childbearing ages to estimate the number of births every year. Each cohort of children born is also followed through time by exposing it to mortality. Finally, the component method takes into account any in-migrants who are incorporated into the population and out-migrants who leave the population. Migrants are added to or subtracted from the population at each specific age. The whole procedure is repeated for each year of the projection period, resulting in the projected population by age and sex, as well as crude death and birth rates, rates of natural increase, and rates of population growth for each year.

The projection can be carried out by single ages or by groups of ages. Although most population projections are currently made by 5-year age groups, projections by single ages are becoming more frequent and may be the dominant type in the near future. Both alternatives can be described in a similar manner and hence, for simplifying the symbols, the single-age projection is described here.

In a given year \( t \), the base-year of the projection, the male and female population at age \( x \) can be represented by \( MP_x^t \) and \( FP_x^t \), respectively. The proportion of persons of a specific age who survive 1 year (the survival ratio as described in chapter III, Mortality), can be represented, for males and females, as \( MS_x^t \) and \( FS_x^t \), respectively. Considering that a person who survives 1 year is also 1 year older, the female population at age \( x+1 \) at year \( t+1 \) is:

\[
FP_{x+1}^{t+1} = FP_x^t \cdot FS_x^t + G_x^t
\]

Where \( G_x^t \) represents the migration component.
Taking into account the migration component can be a complicated task, due to several factors:

(1) Migrants move in both directions, in and out.

(2) Information on migrants refers to their age at the moment of arrival or departure. For example, a person who migrates at the beginning of the year at age 23 may be 24 years old at the end of the year, while another person who migrates later in the year at age 23 is likely to be still 23 years old at the end of the year.

(3) For migrants arriving in the population, mortality has to be taken into account during the projection period but, for those leaving the population, mortality is irrelevant to the projection.

Most computer programs for making population projections take migration into account with the proper formulas. These formulas are not presented here, but they are provided in the documentation (volume II) for the computer program for urban and rural projections (the RUP program presented in appendix VIII-1).

To continue the projection process, the survivorship formula presented above is applied to all ages except the final open-ended age group and the population under age 1 at the end of the year. These ages require special treatment.

For the open-ended age group (for example, ages 85 and over), the formula is:

\[
FP_{85+} = FP_{84+} \cdot FS_{84+} + G_{84+}
\]

To estimate the population under age 1 at the end of year \( t \) (or at the beginning of year \( t+1 \)), first the number of births during year \( t \) has to be calculated. This is done by taking into account the number of women in childbearing ages and a set of age-specific fertility rates. The product of each age-specific fertility rate times the female population at the same age gives the number of births for women at each specific age. Summing the births for each age, the total number of births is obtained.
In symbols:

\[
B^t = \sum_{x=15}^{49} f^t_x \cdot FP^t_x
\]

Where:

- \( B^t \) is the total number of births at time \( t \); and
- \( f^t_x \) is the age-specific fertility rate.

The value \( B^t \) represents births of both sexes. To separate the female from the male births, the proportion of female births to total births is used. This proportion is obtained from vital statistics and fluctuates from about 0.493 to about 0.483, corresponding to sex ratios at birth of 1.03 to 1.07, respectively. Designating this proportion by \( h \), the number of female births during year \( t \) is:

In symbols:

\[
FB^t = h \cdot B^t
\]

Where \( FB^t \) represents female births.

Male births are obtained by subtracting female births from total births. The population at age 0 in year \( t+1 \) is obtained by multiplying the births by the survival ratio corresponding to the interval from birth to age under 1 (\( S_{b,0} \)). For females:

In symbols:

\[
FP^{t+1}_0 = FB^t \cdot FS_{b,0}
\]

Again, if there are migrants under 1 year of age, the migration component must be incorporated.
The same procedure is used for males. Once the population of each sex is projected for a year, the method is repeated for successive years, and the projection by sex and age is obtained for any desired year.

The Bureau of the Census has developed a computer program called RUP for projecting populations. The method used in this program is similar to the procedure described above, except that it does not use survival ratios. Instead, the program estimates the number of cohort deaths and reduces the population accordingly (see appendix VIII-1 and volume II).

Regional projections. The procedure described above is used also for regional projections, provided that information on mortality, fertility, and migration is available for each of the regions. The most important difference in regional projections is that both internal and international migration have to be taken into account. International migration is treated in the same way as in a national projection. Internal migration, on the other hand, requires information on the regions of origin of the in-migrants and regions of destination of the out-migrants. If the projection is made for urban and rural areas, the procedure is straightforward, and several computer programs are available to carry it out (for example, the RUP program presented in appendix VIII-1 and volume II). For a larger number of regions, it is more difficult, but not impossible, to project them all simultaneously (Willekens and Rogers, 1978). Information about flows of migrants by sex, age, place of origin, and place of destination is seldom published. In addition, the computer's capacity may not be large enough if the number of regions is large.

Regional population projections, when summed to obtain the population for the whole country, may produce some inconsistent trends of mortality and fertility at the national level. To avoid this, it has been suggested to make first a population projection for the whole country to serve as a control total for the sum of the regions. Arguments have been presented both in favor of and against this procedure. Arguments for a control total contend that information for the whole country is frequently of better quality than information for each of the regions because vital events may be recorded by place of registration rather than by place of occurrence. Such misplacement of vital events may result in a distorted estimate of the components of growth of each region, and hence, their sum may not reflect the proper total for the country. An argument against a control total is that, if vital registration is reliable, whatever happens in a country will be the result of what happens in each of the regions.

For the few countries that produce regional population projections, there is usually a projection for the whole country serving as a control, and the regional projections are adjusted to this national total. It is advantageous to compare the sum of the regional projections with the total derived independently for the whole country. A small difference produces confidence in the regional projections in relation to what is expected for the whole country, while a large difference indicates that there were inconsistencies between the assumptions made for the regional projections and those made for the national projections. The latter situation calls for a revision of the assumptions. Once revised projections result in small
differences, it is recommended to adjust (usually proportionally) the regional projections to the national ones.

For developing countries where information on interregional migration flows is not available, regional projections still can be produced by using *net* migration flows. As described in chapter V on Migration, net intercensal migration can be estimated for regions if there are two censuses 10 years apart. If, in addition, mortality and fertility can be estimated for each region based on vital registration data or indirectly from census data, then it is quite feasible to make regional population projections. In this case, a comparison of the sum of the regions with the country total is a requirement, and the adjustment of the regional projections to the country total is also highly recommended. See contingency table in appendix VIII-2 for making such an adjustment.

**Base population by age and sex**

A component projection requires a population properly distributed by sex and age to serve as the base population for the starting date of the projection. Reliable estimates of the levels of mortality, fertility and migration are required for the same year. Usually, the base population is taken from the latest available census. But, as discussed in chapter II on Age and Sex Composition, census enumerations are not perfect, and the reported data on the population age and sex structure may be affected by underenumeration of persons in certain ages as well as by age misreporting. During the first years of the projection period, errors in the age and sex composition of the base population may have a large impact on the projected population. As noted above, the component method projects the base population by following cohorts of persons throughout their lifetimes. Thus, if the projection starts with errors in the base year, such errors will be carried throughout the projection period and will have an impact on the projected number of births as well. Suppose, for example, that children age 0 to 4 years were undercounted in the base population. In the projection, not only would the surviving cohorts of these children be smaller than they should be, but when the female cohorts reach reproductive ages, the number of births they have would also be underestimated.

Consequently, before accepting a population to serve as a base for the projections, an evaluation of the completeness of enumeration and the extent of age misreporting should be undertaken and adjustments made as required. Some of the techniques outlined in chapter II can be used for this purpose. Information from post-enumeration surveys also will help in evaluating the quality of the base population data.

In addition to applying the smoothing procedures suggested for adjusting the age structure of the population for possible age misreporting, an evaluation of the enumerated population under age 5 (or under age 10) should not be overlooked. These two age groups, 0 to 4 years and 5 to 9 years of age, frequently have larger errors than other ages, with the possible exception of the very old. Errors in the very old ages are not as important because the
number of persons is relatively small, and these older cohorts will disappear through mortality during the first decades of the projection period. Errors in the younger ages have a more significant impact on the total projection, and they should be corrected. In particular, the younger age groups should be checked for consistency with the estimated levels of fertility and mortality. One feasible procedure for evaluating the population under age 5 years is as follows:

1. Adjust the total population of each sex for estimated undercount.
2. Smooth the population age distribution of each sex if the analysis of the age structure suggests a need for it.
3. Rejuvenate the female population 20 to 59 years of age for the two 5-year periods prior to the census date to represent the female population in reproductive ages 5 and 10 years earlier.
4. Estimate the male and female births during the two 5-year periods prior to the census date.
5. Project the two 5-year birth cohorts to the census date. These figures represent the adjusted population ages 0 to 4 years and 5 to 9 years at the census date.

The computer program BASEPOP is available to carry out these steps. It smooths the population and estimates the population under age 10 years based on the levels of mortality and fertility 10 years prior to the census date. The resulting population is frequently accepted as the base population for the projection. Although the program estimates the population in the age group 5 to 9 years, this population can be replaced by the census population if the user prefers. In the latter case, the population over age 10 years is slightly smoothed (maintaining the enumerated population within each 10-year age group), accepting the census population for ages 5 to 9 years and adjusting only the population under age 5 years. See appendix VIII-3 and spreadsheet documentation in volume II.

A second program (BPSTRNG) adjusts the population under age 10 years using the same procedure as the BASEPOP program, but it makes a strong smoothing of the population over age 10 years. The BPSTRNG program is recommended only for populations with severe age misreporting and no heavy migration. See examples in figure VIII-1 and volume II for documentation of the spreadsheet.

Projection of mortality

There are two steps in projecting mortality: (a) to project the general level of mortality measured by life expectancies at birth; and (b) to estimate the pattern of mortality by age for each of the projected life expectancies.
Figure VIII-1. Age-Sex Pyramids: Actual and Smoothed Populations

Age Group


Male Female

Thousands

Source: U.S. Bureau of the Census.
Projecting the level of mortality. There are two steps in projecting the level of mortality: (a) to assign a target level for a specific date in the future; and (b) to determine a trend of mortality between the base-year level and the target level.

Assigning a target level. The method of estimating a future level of mortality depends on the information available. If past estimates of life expectancy are available, they can be used to determine an expected trajectory for the near future. For this purpose, a logistic function can be fitted to the information if estimates are available for more than one date. The fitting of the logistic is recommended because this function approximates the expected changes in life expectancy at birth. In countries where mortality has already declined from high levels to the lowest levels ever observed, trends indicate that life expectancy increases rather slowly when mortality is very high (life expectancies of only 30 to 35 years), accelerates at middle levels (life expectancies around 50 to 55 years), and increases again rather slowly at low levels of mortality (life expectancies over 65 years). The logistic function has a similar pattern. Two spreadsheets (FITLGSTC and EOLGST) are available for fitting a logistic to historical information (see figure VIII-2 and table VIII-1). The first program requires data for three equidistant dates, or a multiple of three, and does not require asymptotic values. The second one does require asymptotic values, but there are no restrictions on the number or spacing of observations as long as there is more than one. See appendix VIII-4 and spreadsheet documentation in volume II.

Life expectancy levels extrapolated with a logistic function may increase faster or more slowly than the expected future trend of mortality in a particular country. For this reason, it is recommended to evaluate the results after fitting the logistic function to decide whether they yield an acceptable future target for the individual country circumstances. For example, information concerning possible programs of public health should be taken into account in judging the results.

In cases where past estimates of life expectancies cannot be used or do not exist, the method described above does not apply. For such circumstances, life expectancies can be projected based on increases related to the general level of mortality. Such recommended increases have been developed by the United Nations (1989) by analyzing trends in life expectancies for countries with available data and by using logistic functions.

Determining a trend. Once a forecast of the level of life expectancy has been set for a specific future year, the next step is to estimate a trend in life expectancy for the period between the latest available estimate (usually pertaining to the base year of the projections) and the one selected for the future. A logistic function can again be used. This time it is fitted using the base-year value and the target value, along with two asymptotes (Arriaga, 1984), to determine the intermediate levels of life expectancy at birth (see figure VIII-3). The microcomputer program EOLGST is available for this purpose. (See appendix VIII-4 and documentation of the spreadsheet in volume II.)
Figure VIII-2. Actual and Logistically Adjusted Life Expectancies at Birth: Mexico and Japan

Source: U.S. Bureau of the Census.
### Table VIII-1. Actual and Logistically Adjusted Life Expectancies at Birth: Mexico and Japan, Selected Years

<table>
<thead>
<tr>
<th>Year</th>
<th>Male</th>
<th>Female</th>
<th>Actual Male</th>
<th>Actual Female</th>
<th>Adjusted Male</th>
<th>Adjusted Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>1895</td>
<td>24.3</td>
<td>24.5</td>
<td>24.85</td>
<td>25.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>25</td>
<td>25.6</td>
<td>25.17</td>
<td>25.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1905</td>
<td>25.62</td>
<td>26.23</td>
<td>26.23</td>
<td>27.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1910</td>
<td>27.3</td>
<td>27.9</td>
<td>26.23</td>
<td>27.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1915</td>
<td>27.05</td>
<td>28.05</td>
<td>28.05</td>
<td>28.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920</td>
<td>28.15</td>
<td>29.42</td>
<td>28.15</td>
<td>29.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1921</td>
<td>33.7</td>
<td>35.6</td>
<td>28.41</td>
<td>29.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1925</td>
<td>33</td>
<td>34.7</td>
<td>31.49</td>
<td>33.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1930</td>
<td>33.89</td>
<td>36.21</td>
<td>33.89</td>
<td>36.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1935</td>
<td>37.7</td>
<td>39.8</td>
<td>36.85</td>
<td>39.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940</td>
<td>40.38</td>
<td>43.19</td>
<td>40.38</td>
<td>43.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1945</td>
<td>46.2</td>
<td>49</td>
<td>44.4</td>
<td>47.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>54.4</td>
<td>59.6</td>
<td>53.3</td>
<td>57.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>58.63</td>
<td>62.74</td>
<td>61.48</td>
<td>66.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td>65.47</td>
<td>70.12</td>
<td>65.47</td>
<td>70.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td>68.55</td>
<td>73.37</td>
<td>68.55</td>
<td>73.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Male</th>
<th>Female</th>
<th>Actual Male</th>
<th>Actual Female</th>
<th>Adjusted Male</th>
<th>Adjusted Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>1935</td>
<td>46.92</td>
<td>49.63</td>
<td>45.69</td>
<td>48.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940</td>
<td>49.42</td>
<td>52.64</td>
<td>49.42</td>
<td>52.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1945</td>
<td>53.44</td>
<td>57.08</td>
<td>53.44</td>
<td>57.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1947</td>
<td>50.06</td>
<td>53.96</td>
<td>50.06</td>
<td>53.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>57.37</td>
<td>61.41</td>
<td>57.37</td>
<td>61.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>63.6</td>
<td>67.75</td>
<td>61.07</td>
<td>65.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td>65.32</td>
<td>70.19</td>
<td>64.41</td>
<td>69.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td>67.74</td>
<td>72.92</td>
<td>67.33</td>
<td>72.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td>69.31</td>
<td>76.66</td>
<td>69.79</td>
<td>76.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td>71.73</td>
<td>76.89</td>
<td>71.83</td>
<td>77.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>73.35</td>
<td>78.76</td>
<td>73.46</td>
<td>78.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>74.78</td>
<td>80.48</td>
<td>74.76</td>
<td>80.39</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Mexico 1895 to 1960 from Eduardo Arriga, 1968, New Life Tables for Latin American Populations in the Nineteenth and Twentieth Centuries, Population Monograph Series, No. 3, University of California, Berkeley; other information from U.S. Bureau of the Census.
Figure VIII-3. Actual and Logistically Extrapolated Life Expectancies at Birth: Mexico and Japan

Source: U.S. Bureau of the Census.
Estimating the age pattern of mortality. A particular age pattern of mortality pertains to only one level of life expectancy at birth. The opposite is not true: a particular level of life expectancy at birth can be reproduced by various age patterns of mortality. Hence, the problem is to determine which of all the possible age patterns should be used for each of the projected levels of life expectancy. Several procedures are possible; the ones most frequently used are presented here.

Model mortality pattern. One procedure for obtaining a pattern of mortality pertaining to each level of life expectancy at birth is to adopt a model pattern. To do this, a model life table is constructed corresponding to each specified life expectancy at birth. Any model life table can be used for this purpose, but those developed by Coale and Demeny and by the United Nations are the most frequently used because they are available on microcomputer programs. Once model life tables have been generated for each mortality level for each sex, they are used in the population projection. This procedure assumes that the age pattern of mortality of the population is the same as the age pattern of the model. It is up to the researcher to judge whether or not this assumption is acceptable in a particular case.

For the few countries or areas where female life expectancy at birth is still lower than that for males, model patterns may produce an erroneous relationship between the age patterns of mortality by sex. The result may show significantly higher mortality for females than males at all ages. In actuality, in most countries or areas where life expectancy at birth for females is lower than that for males, infant mortality for males is higher than or similar to that for females, and mortality at the older ages (over 50 years) is also higher for males than for females. In other words, in these countries, female mortality is usually higher than male mortality only at ages 1 to 50 years, and the use of model life tables will not produce acceptable results. For an example, see appendix VIII-5.

Change derived from model mortality patterns. Another possibility is to use the change of the age pattern of mortality from model life tables to project change based on an empirical pattern of mortality. For instance, suppose there is an empirical life table for a population with a life expectancy at birth of 60 years, and a mortality pattern for a level of life expectancy of 65 years has to be estimated. The following procedure can be used:

1. Generate two model life tables, with life expectancies of 60 years and 65 years, respectively.
2. Calculate the relative change in the central death rates between the two model life tables.
3. Apply the change calculated in (2) to the central death rates of the empirical life table (see figure VIII-4A).

The United Nations developed a similar procedure using logits. It is available in a microcomputer program called MATCH in the MORTPAK package (see figure VIII-4B).
Figure VIII-4. Empirical and Projected Mortality Rates

Central death rates

Note: Empirical and projected rates pertain to a life expectancy of 58.5 years and 69.8 years, respectively.

Source: U.S. Bureau of the Census.

Probability of dying

Note: Empirical and projected rates pertain to a life expectancy of 64.7 years and 73.7 years, respectively.

Source: U.S. Bureau of the Census.
The procedure of applying changes in model mortality patterns to empirical data is applied separately for each sex. When the future pattern of mortality for each sex is to be projected from empirical data at very low levels of life expectancy (corresponding to high levels of mortality), the resulting relationship between male and female infant mortality can be completely distorted. For instance, suppose there is a population with life expectancies at birth for males and females of 47.6 years and 48.3 years, respectively, with corresponding infant mortality rates for males and females of 137 and 124 per 1,000 live births. Suppose further that the levels of life expectancy at birth for several decades into the future have been forecasted at 69.5 years and 75.8 years for males and females, respectively. If the mortality pattern were estimated by calculating the change between selected model life tables (Coale-Demeny, West region) and applying the change to the empirically derived life table for the base year, the resulting infant mortality rates would be 33 and 15 per 1,000 live births for males and females, respectively. The sex differential of the projected infant mortality rates seems to be too large. Similar inconsistencies in other ages also may be produced when this procedure is used. However, they will appear only if the forecast implies a large reduction of mortality during the projection period or if mortality in the base year was already at a low level. See examples for Mexico and Nepal in appendix VIII-5.

**Interpolation procedure.** An interpolation procedure is suggested for avoiding inconsistencies. The procedure provides a control over the sex differentials of mortality and avoids inconsistencies in the expected pattern of mortality by age. The procedure is carried out as follows: having two life tables with different life expectancies at birth, the probabilities of dying ($q_x$ values) are interpolated between the two life tables in an attempt to derive a new life table with the desired life expectancy at birth. As the initial interpolated values are not likely to produce such a life table, a series of successive iterations is generated to obtain a set of probabilities of dying that will reproduce the desired life expectancy at birth.

The two life tables used for the interpolation process are: (a) an empirical life table corresponding to the base year of the projection; and (b) an ultimate life table with an adequate level of mortality for the interpolation process. The Bureau of the Census has developed two spreadsheets for the purposes of interpolating male and female life tables: INTPLTM and INTPLTF, respectively. Each program has an ultimate life table: one for males and one for females. They were derived by taking the lowest rates specific by age, sex and cause of death (groups of causes) among Western European countries, the United States, Canada, Australia, New Zealand, and Japan during the period 1970 to 1988. The life expectancies at birth for these life tables are just over 81 years and 87 years, for males and females, respectively (see figure VIII-5 and table A-VIII-5.3 of appendix VIII-5).

The advantage of this interpolation process is that it provides not only a better control over the pattern of the age-specific death rates, but also the possibility of easily changing the ultimate life table whenever new information becomes available.
The interpolation process suggested above may have other uses as well. For particular areas of some countries, it is expected that mortality may increase in the future due to the AIDS epidemic. The pattern of mortality under such an epidemic will be different from current standards. Mortality of infants will probably increase together with mortality of young adults. However, mortality of children between 5 and 15 years of age, as well as that of persons over 55 years of age, may continue to decline. No existing models produce changes such as these in the mortality pattern, but AIDS models currently being developed may provide one. In the interactive interpolation procedure, such a model could be substituted for the "upper limit life table" to derive the required intermediate values.
Projection of fertility

The projection of fertility shares certain similarities with the projection of mortality. In most cases, the level of the total fertility rates is projected first, and then the pattern of the age-specific fertility rates is estimated. Occasionally, these steps are reversed: the age pattern of fertility is projected first, and then the corresponding total fertility rates are calculated. For developing countries, the first sequence is the most frequently used because the second one requires reliable information and historical time series that only developed countries are likely to have. Both procedures are acceptable and are subject to producing forecasts that fail to correspond to actual values as measured at some future date. These procedures are discussed in the following sections.

Projecting the level of fertility first. As in the case of projecting the level of mortality, there are two steps in projecting the level of total fertility rates: (a) to assign a target level for a specific date in the future; and (b) to determine a trend of total fertility rates between the base-year level and the target level.

Assigning a target level. The procedure for forecasting a fertility level depends on the available information concerning not only fertility but also other characteristics of the population. Various characteristics have an impact on fertility, and so an analysis of the observed trends in these characteristics could help to determine whether or not fertility will change in the future and the probable pace of any change. For example, an increase in age at marriage would reduce the proportion of married women, and hence fertility may decline. An increase in the proportion of women using contraceptives, or in women’s educational attainment, or in the labor force participation of women in the modern economic sector, also would tend to reduce fertility. The magnitude of fertility reduction cannot be established exactly, but various studies have shown that there is a relationship between such characteristics and the level of fertility (United Nations, 1984).

Attempts have been made to use correlation techniques to establish rather simplistic relationships between crude birth rates and the proportion of women using contraceptives (Nortman and Hofstatter, 1978). Also, procedures have been developed to analyze relationships between total fertility rates and the proportion of married women in childbearing ages, proportion of women using contraceptives, effectiveness of contraceptive methods used, and other variables (Bongaarts and Kirmeyer, 1980). To apply these procedures, it is necessary to measure the levels of the variables used and to forecast them into the future. Often the primary purpose of applying these procedures has been to estimate the number of women that would have to be using contraceptives to achieve specified future levels of total fertility rates.

If records about past levels of total fertility rates exist for a population, a fertility trend can be estimated and projected into the future. Currently, fertility is declining in most countries of the world, except
those which already have achieved very low levels of total fertility rates. Hence, to forecast the fertility level in developing countries is usually to estimate the pace of fertility decline. In a few cases, where fertility has not yet begun to decline, the task is to estimate when a decline will begin.

Based on historical information, one observes that total fertility rates begin to decline rather slowly from high levels (total fertility rates around 6 or 7 births per woman, or even higher); once the decline has started, it accelerates; and then the pace eventually begins to slacken when fertility achieves rather low levels (total fertility rates around 3 births per woman). This pattern of change approximates the shape of a logistic function, and hence, this function can be used to determine past trends of fertility and to extrapolate them to any desired date in the future. The values of the observed total fertility rates for selected countries and the corresponding logistic function fitted to them are presented in figure VIII-6 and table VIII-2.

There are two spreadsheets for extrapolating the trend of fertility: TFRLGST and FITLGSTC (see appendix VIII-6 for the first and appendix VIII-4 for the second one). The latter spreadsheet requires information for three equidistant dates, or a multiple of three, and does not require asymptotic values. The first one does require asymptotic values, but there are no restrictions on the number or spacing of observations, as long as there is more than one. The extrapolated value should not be accepted blindly; rather, it should be evaluated to determine its feasibility. Such an evaluation can be based on: (a) other countries' experiences concerning the decline of total fertility rates from levels similar to those in the population being analyzed; or (b) whether or not the forecasted value implies that the country will have plans to promote a change in the level of fertility, and whether the means to achieve the change will be available. If the value extrapolated with the logistic function is not acceptable, a more reasonable forecast should be tried.

**Projecting a trend of fertility.** Once a value of the total fertility rate has been accepted as a forecast for a specified future date, the logistic function can also be used to interpolate between the latest available hard estimate and the value forecasted for the future. The program TFRLGST can again be used for this purpose (see example in figure VIII-7).

There are some populations where fertility has not yet started to decline. In this case, the use of the logistic function to interpolate between the current level of fertility and the one assumed for the future may produce an unacceptable decline during the first years of the interpolation. The fertility decline may start too soon in the interpolated values. In such a case, there is another option: to interpolate the total fertility levels by using a sine function. This function has almost a flat trend at the beginning of the interpolation and then starts to decline. The Bureau of the Census has developed a spreadsheet, TFRSINE, that interpolates the total fertility rates using a sine function. See appendix VIII-6.
Figure VIII-6. Actual and Logistically Adjusted Total Fertility Rates: Taiwan, 1958 to 1987

Source: Statistical Yearbook of the Republic of China, 1988

Figure VIII-7. Actual and Logistically Extrapolated Total Fertility Rates: Mexico, 1960 to 2025

Source: U.S. Bureau of the Census.
Table VIII-2. Actual and Logistically Adjusted Total Fertility Rates
Taiwan, 1958 to 1987

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual</th>
<th>Adjusted</th>
<th>Year</th>
<th>Actual</th>
<th>Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958</td>
<td>6.06</td>
<td>6.08</td>
<td>1973</td>
<td>3.21</td>
<td>3.32</td>
</tr>
<tr>
<td>1959</td>
<td>5.99</td>
<td>5.95</td>
<td>1974</td>
<td>3.05</td>
<td>3.14</td>
</tr>
<tr>
<td>1960</td>
<td>5.75</td>
<td>5.81</td>
<td>1975</td>
<td>2.83</td>
<td>2.98</td>
</tr>
<tr>
<td>1961</td>
<td>5.59</td>
<td>5.66</td>
<td>1976</td>
<td>3.08</td>
<td>2.82</td>
</tr>
<tr>
<td>1962</td>
<td>5.47</td>
<td>5.49</td>
<td>1977</td>
<td>2.70</td>
<td>2.68</td>
</tr>
<tr>
<td>1963</td>
<td>5.35</td>
<td>5.32</td>
<td>1978</td>
<td>2.71</td>
<td>2.54</td>
</tr>
<tr>
<td>1964</td>
<td>5.10</td>
<td>5.13</td>
<td>1979</td>
<td>2.66</td>
<td>2.42</td>
</tr>
<tr>
<td>1965</td>
<td>4.83</td>
<td>4.94</td>
<td>1980</td>
<td>2.52</td>
<td>2.31</td>
</tr>
<tr>
<td>1966</td>
<td>4.82</td>
<td>4.74</td>
<td>1981</td>
<td>2.46</td>
<td>2.22</td>
</tr>
<tr>
<td>1968</td>
<td>4.33</td>
<td>4.32</td>
<td>1983</td>
<td>2.16</td>
<td>2.05</td>
</tr>
<tr>
<td>1969</td>
<td>4.12</td>
<td>4.11</td>
<td>1984</td>
<td>2.05</td>
<td>1.98</td>
</tr>
<tr>
<td>1970</td>
<td>4.00</td>
<td>3.91</td>
<td>1985</td>
<td>1.89</td>
<td>1.92</td>
</tr>
<tr>
<td>1971</td>
<td>3.71</td>
<td>3.71</td>
<td>1986</td>
<td>1.68</td>
<td>1.86</td>
</tr>
<tr>
<td>1972</td>
<td>3.37</td>
<td>3.51</td>
<td>1987</td>
<td>1.70</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Adjusted rates: U.S. Bureau of the census. The adjusted logistic function has asymptotes of 7.1 and 1.5.

The last step in projecting fertility is to choose from the results of the logistic the total fertility rates to be used as input to the population projection. In most projections with declining fertility, total fertility rates are specified for every 5 years. In population projections for countries where fertility has already achieved very low levels (below replacement level, for example), such detailed input is not required. Once fertility reaches such low levels, only small fluctuations in total fertility rates are expected, and input values for every 5 years would be superfluous.

**Estimating future patterns of fertility.** The relationship of age-specific fertility rates to the level of the total fertility rate is analogous to the relationship of patterns to levels of mortality. While a particular set of age-specific fertility rates corresponds to only one value of total fertility rate, a given total fertility rate can be reproduced by an infinite number of sets of age-specific fertility rates. Hence, the task is to determine which set of age-specific rates should be chosen to correspond to each total fertility rate to be used in a population projection. In real populations, the proportional distribution of the age-specific fertility rates may change as the level of fertility changes. Therefore, the selection of an age pattern of
fertility should not be overlooked, as the same total fertility rate with a different age pattern would result in different numbers of births, different crude birth rates, and different rates of natural increase throughout the projection period.

Models have been developed for estimating the pattern of age-specific fertility rates based on the level of contraceptive prevalence, age at marriage, and the pace of women entering marriage (Coale and Trussell, 1974). These models are useful for other purposes, but it usually is not feasible to incorporate them into population projections because they require measures of contraceptive prevalence and patterns of marriage for women that are difficult to project into the future.

Actually, the selection of a pattern of age-specific fertility rates for each projected level of total fertility rate is fairly straightforward in most cases. Frequently, a pattern of age-specific fertility rates is adopted from other countries that have achieved the specified level of fertility and that have characteristics similar to those of the country whose population is being projected. The condition of similarity is not easy to define nor to find among the countries of the world. In general, more importance is given to the level of fertility than to the characteristics of the population. A microcomputer program (ASFRPATT) is available for selecting age-specific fertility rates pertaining to a given level of total fertility rate (see appendix VIII-7). The program incorporates patterns of fertility rates observed in developing countries. Some developing countries may have a fertility pattern different from those included in the program, but new patterns easily can be substituted to accommodate such situations. The patterns currently included in the program are presented in figure VIII-8.

Projecting the age pattern of fertility first. As noted at the outset, the projection of fertility can also be made by reversing the order of the steps described above, that is, by forecasting the age-specific fertility rates first, and then calculating the corresponding total fertility rates. This technique is useful for countries with reliable historical information on age-specific fertility rates. Such historical data allow the analysis of fertility using period fertility rates as well as cumulative fertility rates by age of female cohorts from the past to the present. The trends of fertility change and the differential fertility levels among cohorts in the same year and among different cohorts at the same age are taken into account. The forecast of the rates (either age-specific rates or cumulative rates) is made by using correlation analysis or by fitting mathematical functions to the data. Once age-specific fertility rates have been either projected directly or calculated from the projected cumulative rates, the corresponding total fertility rates are derived.

These procedures have been applied in developed countries using age-specific fertility rates for 5-year age groups and by single ages and single calendar years (Das Gupta, 1989). Because developed countries already have low fertility, future fluctuations are expected to be small. For developing countries, on the other hand, the technique is rarely applicable because fertility fluctuates more considerably, and reliable historical information is seldom available.
Projection of migration

Migration is the third component of population change. Chapter V on Migration referred to the two main problems faced in analyzing migration: (a) the measurement of population moves; and (b) the paucity of migration data available for most countries. These problems pertain to both internal and international migration, making the predictability of future trends of migration more uncertain than projections of mortality and fertility.

Internal migration may be more predictable than international migration. Within a country, people are motivated to move primarily by social and economic factors. Once these factors start to operate in a region, their effect continues to be felt for several years. Regional development plans are put into place, and emerging large cities become places of attraction for persons living in the less developed areas of the country.

International migration, on the other hand, may be a result not only of economic conditions and differentials among countries but also of political unrest, persecutions, famines, and other extreme conditions in the countries.
of origin. Thus, international migrants may not only feel rejected by stagnated economies and attracted by industrialized societies, but may flee in large numbers as refugees looking for better lives elsewhere.

Due to the unpredictability of conditions such as crop failure, emerging violence, and bellicose activities in any country of the world, migration forecasts are subject to large errors. These errors, as well as the migration itself, may have different significance in relation to the population of the area of origin and that of the area of destination. For example, in the international context, the number of migrants to India may not be significant in relation to India's population, but an equal number of migrants going to Djibouti would have a large impact on that nation's small population. While the number of Pakistanis going to Saudi Arabia constitutes a small proportion of Pakistan's population, it constitutes a large proportion of the labor force of the receiving Arab country. In contrast, the migration of persons from Yemen to Saudi Arabia has a significant impact in the population of both countries.

In the case of internal migration, several examples come to mind. In the 1950's, the transfer of Brazil's national government to Brasilia created some congestion and disorder that had to be dealt with in the planned city as it received its new inhabitants. Migration to the capital and the concomitant problems of providing housing and services to the burgeoning population continue even today, yet the migration to Brasilia has little impact on the rest of the country. Likewise, in Venezuela in the 1960's and 1970's, Ciudad Guyana was an area of migrant attraction as a consequence of a planned regional economic development. This area received many migrants from other regions of Venezuela, but again, the migration had little demographic impact on other regions of the country. Finally, in more recent years, the movement of Mexico's national statistical office from Mexico City to Aguascalientes caused a housing shortage and other difficulties at the new site as the employees and their families settled in their new location, but the move had little effect on the population that remained in Mexico City.

Not only the total level of migration, but also the age and sex composition of migrants is difficult to forecast. If a move is motivated by economic factors, migrants are likely to be concentrated in the young adult ages; however, their gender is not easy to predict. In the case of internal migration, the predominant gender of migrants depends not only on the characteristics of the society but also on the historical moment in the country. In the case of international labor migration, migrants are likely to be predominantly male. The sex and age composition even of refugees can fluctuate widely, depending on the reasons for the move and the characteristics of the society affected. The age and sex composition of refugees often depends on whether they migrate as individuals or as part of a family unit. Finally, the number of child migrants depends principally on the marital composition of adult female migrants.

Both internal and international migration can be only partially controlled, although recently many countries affected by international migration are intensifying enforcement of existing laws in an attempt to gain a more stringent control over persons crossing their borders. Internal
migration can also be controlled, but only by a well-organized and well-administrated system, often requiring the use of compulsion.

Facing the uncertainty of forecasting migration, only general guidance will be presented here. Internal and international migration are considered separately.

Internal migration. Past information on internal migration is needed to make a reasonable forecast of the number of persons that probably will change residence in the future. The base information required depends on the number of areas for which the population is to be projected. In most cases, only urban and rural areas are projected, but state and/or regional projections are increasingly in demand for national planning purposes, and many countries will undoubtedly have to produce them in the near future.

The estimation of past internal migration is required not only for estimating the total number of migrants, but also for calculating migration rates by age and sex based on the population of the area of origin (see chapter V on Migration). The next step is to decide whether or not to forecast a change in the rates for the future and, if so, to what levels. Rates of migration from rural to urban areas tend to increase during the process of development; however, once a country's population achieves a high proportion in urban areas, rural out-migration rates tend to decline slightly and then to stabilize. Historical information on these rates is the best indicator of possible future trends. In developing countries, the rural population usually is still increasing. In this situation, maintaining a constant level of out-migration rates increases the actual number of migrants from rural to urban areas because the constant rates are applied to an ever-larger base population. If past information on migration rates by age and sex is not available, some patterns of rates are available from migration models (see Rogers and Castro, 1981). However, the question of determining the levels corresponding to these patterns still persists. To project the level of migration, national plans for the future should be taken into account, for example, plans for improving communication and transportation systems, development programs for certain areas of the country, and information systems concerning regional availability of job opportunities.

In cases where population projections are to be made for various regions simultaneously, special information is required on migration rates by age and sex, by place of origin and place of destination. Forecasting such rates is rather cumbersome, and the most frequent assumption is that they will remain constant in the future. However, if the population of each region is to be projected independently, there are two possibilities: (a) to use net migration rates by age and sex calculated based on the regional population; or (b) to use the absolute number of migrants. For regions where net migration is positive (that is, migrants coming in are more numerous than those leaving the region), migration rates calculated based on the region's total population may produce unexpected results. Since the region is an area of immigration, even if the projection assumes constant migration rates, the absolute number of migrants may increase rapidly because of the growth of the region's total population. For this reason, it is sometimes preferable to use absolute
numbers of migrants rather than migration rates when projections are made independently for each region.

**International migration.** As mentioned above, the forecast of international migration is one of the most difficult to make. Only those countries that are considered by migrants to be of greatest attraction may have the possibility of forecasting migration with a certain accuracy, by offering only a specific number of immigrant visas. Even then, the issue of undocumented migrants remains unresolved.

In projecting the population of most countries of the world, it is assumed that international migration will be nil. For those populations where international migration is significant, the estimated number of migrants during the past is frequently held constant in projecting to the future unless there are indications that the volume of migration may change. In almost all cases, because of the uncertainty of international migration in the distant future, it is assumed that by the end of the projection period international migration will be nil, particularly if the projection is made for a period longer than 30 or 40 years.

Finally, as noted above, the age and sex composition of international migrants depends on the situation in each country. If information is not available, model patterns by age and sex can be used.

**Comments on component projections**

As mentioned above (see section on base population), errors in census populations require an analysis of the base population to be used in a projection. Errors in the base population will have an impact on the projection. For example, the census and strongly smoothed populations presented in figure VIII-1 for the year 1982 were each projected with the same levels of mortality and fertility, and without migration. The impacts of the errors of age misreporting on the population projections for the years 1992, 2005 and 2020, respectively) are presented in figure VIII-9, showing that the errors in the base population are evident in the cohorts throughout the projection period. The population projection is presented in appendix VIII-8.

The uncertainty in forecasts of the components of population growth, as well as errors that may exist in the population used as a base for the projection, suggest a need for making more than one assumption regarding one or more of the components of population growth. For developing countries, where fertility is still high, the projected fertility levels are subject to errors. Hence, more than one assumption is usually made, representing high, medium, and low fertility levels in the future.

Mortality can be forecast with more confidence: in most population projections, only one mortality assumption is made. However, with an increase in the AIDS epidemic in many countries, it may be useful to begin soon to include more than one mortality forecast. Also, an increasing interest in the aging of the population in developed countries may call for future projections with more than one mortality hypothesis.
Figure VIII-9. Projected Census and Smoothed Population by Age and Sex, Selected Years

Source: U.S. Bureau of the Census.
Population projections for most countries include only one assumption about future international migration, although this variable has a high degree of uncertainty. Some countries, for example, Australia and the United States, use more than one hypothesis concerning migration. In most developed countries, given the low levels of fertility and mortality, international migration is becoming the most important component of population change. Hence, the number of population projections with more than one migration hypothesis is likely to increase in the near future.

For countries where international migration is significant, population projections for regions within the country have to determine which regions will receive the immigrants and which will provide the emigrants.

Finally, population projections should be updated regularly as new information becomes available. Frequent updating is becoming more and more feasible with the proliferation of microcomputers.

Projection Methods Based on Mathematical Functions

Mathematical methods are applied mainly for projecting the population of small areas within a country; they are seldom used for projecting the population of the country as a whole. These methods do not provide the age structure of the population. Although the total for each sex can be obtained, projections for each sex are made independently and thus are not necessarily consistent. Nevertheless, the technique is useful for projecting the population of small areas without vital statistics.

Several mathematical functions can be used for projecting the total population; the most frequently used are linear, exponential, and logistic functions. Such projections are usually made by extrapolating past population trends into the future. When such a procedure is applied to the population of the whole country, the projections may be subject to errors due to lack of control over changes in the components of population growth and age structure. For small areas, the results are more acceptable because the projections can be constrained by an overall projection pertaining to an aggregated total of the small areas. For example, a country with a population projection prepared by the component method at the state level can use a mathematical function to project the population of small areas within each state under the condition that the sum of the populations of the small areas will equal the state projection. Once the mathematical projection for each small area has been made, the population of each area can be adjusted proportionally to the total population of the state. Thus, the state population serves as a control total for the projections of the small areas.

Linear function

One of the simplest ways to project the population of an area is to assume that the average annual increase that occurred during a recent period of time will be repeated in the future. For instance, if a population grew from 1,254 in 1972 to 2,150 in 1979, the annual average increase is the difference between the two figures divided by the number of years in the
interval: \((2,150 - 1,254) / 7\), or 128 per year. To calculate a linear projection for the year 1990, the annual increase times the number of years between 1979 and 1990 is added to the 1979 figure: \(2,150 + (11 \times 128) = 3,558\).

In symbols:

\[ p_{t+n} = p_t + n \cdot Y \]

Where:

- \(p_t\) represents the population at year \(t\);
- \(n\) is the number of years to project from year \(t\); and
- \(Y\) is the average annual increase of the population.

The average annual increase is calculated as follows:

\[ Y = (p_t - p^0) / t \]

Based on population information for years 0 and \(t\).

Rates can also be used in the linear projection.

**Exponential function**

Under conditions of constant mortality and fertility, and in the absence of migration, human populations tend to grow exponentially. Thus, it has been suggested that an exponential rather than a linear function would better represent the growth of a population. The formula used for an exponential projection is:

In symbols:

\[ p_{t+n} = p_t e^{nt} \]

Where the symbols represent the same concepts as above, but the annual growth rate is an exponential growth rate, and \(e\) is the base of the neperian logarithm.
If the growth rate is not available, it can be calculated based on the population for two previous dates. For instance, using the same values as before, the average annual exponential growth rate can be calculated as:

In symbols:

\[ r = \frac{\ln (p_{1982} / p_{1973})}{n} \]

Where:

\( \ln \) represents the natural logarithm.

Replacing the values:

\[ r = \frac{\ln (150,348 / 123,876)}{9} = 0.0215 \]

or, an annual average exponential growth rate of 2.15 percent.

Once the average annual exponential growth rate has been calculated, the projection of the population under the assumption of exponential growth is obtained by applying the above formula. For example:

If:

\[ p_{1982} = 150,348 \]
\[ r = 0.0215, \text{ and} \]
\[ n = 17 \text{ years,} \]

then:

\[ p_{1999} = 150,348 \cdot e^{17 \times 0.0215} = 150,348 \cdot 1.44170 \]
\[ p_{1999} = 216,756 \]

Comments on linear and exponential projections

It is well known that when the same growth rates are used in a linear and an exponential extrapolation, estimates based on the exponential function are consistently higher than those based on the linear function. See table VIII-3 and figure VIII-10.
Table VIII-3. Linear and Exponential Projections of Population

Total population in year 1982: 238,765
Rate of population growth (percent): 2.8

Projected Populations

<table>
<thead>
<tr>
<th>Year</th>
<th>Linear</th>
<th>Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>238,765</td>
<td>238,765</td>
</tr>
<tr>
<td>1983</td>
<td>245,450</td>
<td>245,545</td>
</tr>
<tr>
<td>1984</td>
<td>252,136</td>
<td>252,517</td>
</tr>
<tr>
<td>1985</td>
<td>258,821</td>
<td>259,688</td>
</tr>
<tr>
<td>1995</td>
<td>325,675</td>
<td>343,601</td>
</tr>
<tr>
<td>2005</td>
<td>392,530</td>
<td>454,628</td>
</tr>
<tr>
<td>2015</td>
<td>459,384</td>
<td>601,532</td>
</tr>
<tr>
<td>2025</td>
<td>526,238</td>
<td>795,905</td>
</tr>
<tr>
<td>2035</td>
<td>593,092</td>
<td>1,053,086</td>
</tr>
</tbody>
</table>

Figure VIII-10. Linear and Exponential Population Projection

Note: Projected with the same annual growth rate of 3 percent.
In general, it is recommended that projections based on mathematical functions be made for rather short periods of time, not exceeding 15 years. During such short periods, any of the procedures will provide similar results if a projected population figure is available to which the aggregated population of all small areas can be controlled and each one proportionally adjusted.

One further aspect is important to consider. In situations where the populations of small areas have declined, their growth rates will be negative. In such cases, the exponential extrapolation has the advantage over the linear one, in that the projected population figure will never be negative. If the linear assumption is used in these cases, the projected population may become negative if the decline of the population was substantial during the past, or if the projection period is too long.

**Logistic function**

A logistic function also can be used for projecting total populations (Pearl and Reed, 1920). Like the two functions already discussed, the logistic is recommended for projecting small area populations for which no component projections can be made. And again, the projections should not be for long periods of time, and they may be improved if there exists a control population to which the sum of the area populations can be adjusted.

The projection of the total population with a logistic function requires three (or a multiple of three) dates with information, and the dates should be equidistant. This requirement considerably limits the use of logistic functions for the projection of total populations. Nonetheless, if the lower and upper asymptotes of the logistic function are known, then information for only two equidistant dates are required. Although the lower asymptote could be assumed to be 0, there is no reasonable procedure for estimating the upper asymptote, which represents the maximum size of the population in the distant future.

But still the logistic function can be used to project the populations of small areas if a control total exists for the sum of all small areas. In this case, instead of projecting the population of each area, the proportions of the population of each area are projected in relation to the total control population. Since the proportions vary from 0 to 1, the lower and upper asymptotes of the logistic can be assumed to have these values as limits. For instance, suppose a population projection is available for a country by state, and that one of the states has two major cities for which there is no reliable information on vital statistics. The city populations can be projected as follows:

1. Calculate the proportion of each city population to the total state population and the proportion of the non-city population to the total state population for two dates in the past.
2. Fit a logistic function to the pair of proportions pertaining to each area.
(3) Extrapolate the proportions with the logistic function (asymptotes 0 and 1).

(4) Proportionally adjust the projected proportions so their sum is equal to 1.

(5) Multiply the adjusted proportions by the projected population for the state.

Comparison of the three mathematical functions

Examples of the results of the three mathematical functions discussed above are presented in tables VIII-4 and VIII-5. As the data show, differences among the alternatives are rather small when a population control total is available for the aggregated small areas. The example presented in table VIII-4 is based on real data, pertaining to the state populations of Colombia. The projections were made for several states and for the balance of the country. The control population pertains to the actual population for the whole country enumerated in the 1973 and 1985 censuses. The projections were made for 1973 and 1985 based on data pertaining to the 1951 and 1964 censuses. For comparative purposes, the table provides not only the estimated populations for 1973 and 1985 based on information from the previous two censuses but also the actual populations enumerated in 1973 and 1985. The reader may interpret the results of the different functions used. A similar exercise for Ghana is presented in table VIII-5.
Table VIII-4. Example of Projected Populations for Selected Colombian States, Using Three Functions: Linear, Exponential and Logistic.

<table>
<thead>
<tr>
<th>States</th>
<th>1951</th>
<th>1964</th>
<th>1973</th>
<th>1985</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antioquia</td>
<td>1,543,734</td>
<td>2,477,299</td>
<td>3,176,695</td>
<td>3,888,067</td>
</tr>
<tr>
<td>Bogotá, D.F.</td>
<td>715,250</td>
<td>1,697,311</td>
<td>2,861,915</td>
<td>3,982,941</td>
</tr>
<tr>
<td>Cauca</td>
<td>391,905</td>
<td>607,197</td>
<td>716,853</td>
<td>795,838</td>
</tr>
<tr>
<td>Marino</td>
<td>542,763</td>
<td>705,611</td>
<td>802,389</td>
<td>1,019,098</td>
</tr>
<tr>
<td>Santander</td>
<td>747,706</td>
<td>1,001,213</td>
<td>1,233,576</td>
<td>1,438,226</td>
</tr>
<tr>
<td>Valle</td>
<td>1,106,927</td>
<td>1,733,053</td>
<td>2,392,715</td>
<td>2,847,087</td>
</tr>
<tr>
<td>Rest of country</td>
<td>6,914,075</td>
<td>9,262,825</td>
<td>11,597,975</td>
<td>13,896,069</td>
</tr>
<tr>
<td>Total</td>
<td>11,962,360</td>
<td>17,484,509</td>
<td>22,862,118</td>
<td>27,867,326</td>
</tr>
</tbody>
</table>

A) Predictions for years 1973 and 1985 based on information for 1951 and 1964, after adjustment to census totals.

<table>
<thead>
<tr>
<th>States</th>
<th>1973</th>
<th>1985</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antioquia</td>
<td>3,351,510</td>
<td>4,206,093</td>
</tr>
<tr>
<td>Bogotá, D.F.</td>
<td>2,550,638</td>
<td>3,465,584</td>
</tr>
<tr>
<td>Cauca</td>
<td>811,420</td>
<td>1,007,867</td>
</tr>
<tr>
<td>Marino</td>
<td>878,058</td>
<td>1,022,323</td>
</tr>
<tr>
<td>Santander</td>
<td>1,262,570</td>
<td>1,488,857</td>
</tr>
<tr>
<td>Valle</td>
<td>2,324,593</td>
<td>2,896,469</td>
</tr>
<tr>
<td>Rest of country</td>
<td>11,683,328</td>
<td>13,780,113</td>
</tr>
<tr>
<td>Total</td>
<td>22,862,118</td>
<td>27,867,326</td>
</tr>
</tbody>
</table>

B) Differences between census information and predicted values.

<table>
<thead>
<tr>
<th>States</th>
<th>1973</th>
<th>1985</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antioquia</td>
<td>-174,815</td>
<td>-318,026</td>
</tr>
<tr>
<td>Bogotá, D.F.</td>
<td>311,275</td>
<td>517,357</td>
</tr>
<tr>
<td>Cauca</td>
<td>-94,565</td>
<td>-212,029</td>
</tr>
<tr>
<td>Marino</td>
<td>4,331</td>
<td>-3,225</td>
</tr>
<tr>
<td>Santander</td>
<td>-28,994</td>
<td>-50,631</td>
</tr>
<tr>
<td>Valle</td>
<td>68,122</td>
<td>-69,402</td>
</tr>
<tr>
<td>Rest of country</td>
<td>-85,353</td>
<td>115,956</td>
</tr>
<tr>
<td>Total (absolute)</td>
<td>767,455</td>
<td>1,266,626</td>
</tr>
</tbody>
</table>

341
Table VIII-5. Example of Projected Populations for Selected Ghanaian Regions, Using Three Functions: Linear, Exponential and Logistic.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Census population</th>
<th>1960</th>
<th>1970</th>
<th>1984</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accra</td>
<td></td>
<td>491,818</td>
<td>903,447</td>
<td>1,420,066</td>
</tr>
<tr>
<td>Eastern</td>
<td></td>
<td>1,094,196</td>
<td>1,209,828</td>
<td>1,679,483</td>
</tr>
<tr>
<td>Volta</td>
<td></td>
<td>777,285</td>
<td>947,268</td>
<td>1,201,095</td>
</tr>
<tr>
<td>Ashanti</td>
<td></td>
<td>1,108,133</td>
<td>1,481,698</td>
<td>2,089,683</td>
</tr>
<tr>
<td>Rest</td>
<td></td>
<td>3,255,583</td>
<td>4,017,072</td>
<td>5,815,247</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>6,726,815</td>
<td>8,559,313</td>
<td>12,205,574</td>
</tr>
</tbody>
</table>

A) Projections for the year 1984 after adjustment to census totals based on information for 1960 and 1970.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Linear</th>
<th>Exponential</th>
<th>Logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accra</td>
<td>1,623,482</td>
<td>2,087,493</td>
<td>2,044,044</td>
</tr>
<tr>
<td>Eastern</td>
<td>1,504,973</td>
<td>1,373,572</td>
<td>1,374,076</td>
</tr>
<tr>
<td>Volta</td>
<td>1,300,389</td>
<td>1,232,274</td>
<td>1,239,233</td>
</tr>
<tr>
<td>Ashanti</td>
<td>2,199,442</td>
<td>2,194,721</td>
<td>2,206,075</td>
</tr>
<tr>
<td>Rest</td>
<td>5,577,287</td>
<td>5,317,713</td>
<td>5,342,147</td>
</tr>
<tr>
<td>Total</td>
<td>12,205,574</td>
<td>12,205,574</td>
<td>12,205,574</td>
</tr>
</tbody>
</table>

B) Differences between census information and projected values for 1985.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Linear</th>
<th>Exponential</th>
<th>Logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accra</td>
<td>203,416</td>
<td>667,427</td>
<td>623,978</td>
</tr>
<tr>
<td>Eastern</td>
<td>-174,510</td>
<td>-306,111</td>
<td>-305,607</td>
</tr>
<tr>
<td>Volta</td>
<td>99,294</td>
<td>31,179</td>
<td>38,138</td>
</tr>
<tr>
<td>Ashanti</td>
<td>109,759</td>
<td>105,038</td>
<td>116,392</td>
</tr>
<tr>
<td>Rest</td>
<td>-237,960</td>
<td>-497,534</td>
<td>-473,100</td>
</tr>
<tr>
<td>Total absolute differences</td>
<td>824,939</td>
<td>1,607,288</td>
<td>1,557,015</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of the Census.
Appendix VIII-1

Population Projections: the RUP Program

The rural-urban population projection (RUP) program (Johnson, 1990) can be used for projecting the population of two areas simultaneously. In this case, there are two options: (a) to project each of the areas and obtain the total as an aggregate, or (b) to project the total and one area and obtain the other area by subtraction. Additionally, the program can be used to run a projection for the total country only.

The program uses the cohort component method for projecting the population by age and sex. The projection is made following cohorts by single years of age. Using interpolated values of central death rates from life tables, RUP estimates the number of deaths and subtracts them from each cohort. It also takes into account internal and international migration. Every year births are estimated using fertility rates and the female population in reproductive ages. The annual number of births constitute new population cohorts that are followed through in the future (see volume II).

RUP has several distinctive aspects as compared with other frequently used population projection programs. These aspects are:

(1) Updating and projecting the base population. For instance, if a country has a population census in 1983 and has good vital statistics up to 1990, RUP operates as follows: (a) it uses the 1983 base population and the vital statistics from 1983 to 1990 to update the population up to 1990 by age and sex; and (b) then it takes the estimated population in 1990 and projects it to any desired year using the assumed values of mortality, fertility, and migration.

(2) Flexibility of detail of input data. The input data can be compiled in different levels of detail. The most common option is the use of single years of age or 5-year age groups. This detail relates to the assumptions made for the future, as well as to the vital statistics as mentioned above. The detail for each demographic variable is independent; for instance, mortality can be given by single years of age, and migration by 5-year age groups.

(3) Flexibility of reference years of input data. The demographic information used for projecting and/or updating the population can pertain to any year between the year of the base population and the final year of the projection period. As in item (2), each demographic variable can pertain to any year. This allows the user, for instance, to update a population by simulating a catastrophic event such as casualties of famine or war, refugee movements, or fluctuations of fertility during a specific year.
the user specifies. Information on life tables, fertility rates, and internal or international migration used for any year of the projection period can be obtained as output. These detailed results can be provided for the two areas being projected as well as for the third area obtained by addition or subtraction.

The use of the RUP program differs from the use of the spreadsheets presented in this manual. The program was written in FORTRAN language and was compiled for use on microcomputers. To run the program, a file containing input data must be created. Once the file is ready, the program will provide the results. The user can select from several options for the detail desired in the output. See volume II for a detailed explanation of the procedure used for projecting the population in RUP, for creating the input file, and for documentation of the program.
Appendix VIII-2

Contingency Table

Description

This technique adjusts information in a table to a set of desired marginals totals (totals of rows and/or columns). For instance, data classified by two characteristics may need to be adjusted to certain totals different from the sum of each column and row. This technique modifies the original data in such a way that the sums of columns and rows correspond to the desired totals.

Data Required

(1) Original distribution of data to be adjusted.

(2) Desired marginal totals to which the data are to be adjusted.

Assumptions

The technique assumes that the original distribution of the information can be adjusted to obtain an estimate of the actual distribution by successive proportional adjustments of rows and columns through an iterative process.

Procedure

The original distribution of data is adjusted to the desired marginal totals as follows: (a) each column of data is adjusted proportionally to the desired total of the column; (b) each row of data is adjusted proportionally to the desired total of the row; (c) the first two steps are repeated until the data agree with the desired marginal totals.

Advantages

The technique provides an estimated distribution of data that matches any set of desired marginal totals. It is useful as a final step in adjusting population projections by regions or states to the total population of a country.

Limitations

Requires an initial distribution of the data by the characteristics of the marginals.
Software

The Bureau of the Census has developed a spreadsheet to apply this procedure. It is called CTBL32, and its documentation is presented in volume II.
Appendix VIII-3

Technique for Smoothing and Adjusting the Base Population

Description

There are two steps to be followed in preparing the base population for a projection. The first step is to smooth the population ages 10 years and over. The second is to evaluate the population under age 10 years in relation to other age groups. The smoothing process was discussed in chapter II on Age and Sex Composition. Only the second step will be detailed here. To adjust the population under age 10 years to the population in other ages, use the following steps:

1. Rejuvenate the female population in reproductive ages for 5 and 10 years, using the life table function $nL_x$.
2. Estimate the births of the rejuvenated females during the two 5-year periods prior to the census, based on available fertility rates.
3. Separate the births calculated in step (2) into male and female births and project them forward to the census date, using the corresponding mortality from appropriate life tables.

The projected births at the census date represent the adjusted population ages 0 to 4 years and 5 to 9 years.

Data Required

1. Population by 5-year age groups and sex for the base year.
2. The female $nL_x$ function of two life tables pertaining to different dates.
3. The male and female $nL_x$ functions for ages under 1 year, 1 to 4 years, and 5 to 9 years.
4. A set of age-specific fertility rates for the base year.
5. Two estimates of total fertility rates for two different dates in the past.
6. The sex ratio at birth for the base year.

The $nL_x$ function and total fertility rates are interpolated to 2.5 years and 7.5 years prior to the census date. Thus, it is desirable that the life tables and total fertility rates pertain to dates that will allow for an interpolation for these time periods. For instance, if the census was taken in mid 1990 (or 1990.5), then (a) the reference dates of the most recent life
table and total fertility rate should be after the beginning of 1988 (after 1988.0); and (b) the reference dates of the earlier life table and total fertility rate should be before the beginning of 1983 (before 1983.0). It is advisable to adhere to these suggested dates, since the computer program uses them as boundaries to interpolate the values for the rejuvenation process. While reference dates closer than 5 years apart could be used, the researcher should be aware that the program would then extrapolate rather than interpolate, and such results should be taken with caution.

Assumptions

The assumptions concern only migration. If the population has out-migration, the technique assumes that female migrants moved with their children. If the population has in-migration, the technique assumes that the fertility of female migrants is the same as the fertility of resident women.

Procedure

The procedure can be divided into two parts: (a) smoothing the population age 10 years and over; and (b) estimating the population under age 10 years.

Smoothing the population age 10 years and over

There are two procedures for smoothing the population age 10 years and over, as discussed in chapter II: (a) light smoothing if the total population in each 10-year age group is kept as enumerated; and (b) strong smoothing if the totals of each 10-year age group are modified.

Adjusting ages under 10 years

(1) Rejuvenate the female population ages 20 to 59 years (by 5-year age groups) to 5 and 10 years prior to the base year, using the survivorship rates estimated from the $n_l$ life table function previously interpolated to appropriate years. This provides the female population in reproductive ages 5 years and 10 years prior to the base year.

(2) Use the information from step (1) and the base population to obtain an average of the female population in reproductive ages (15 to 49 years, by 5-year age groups) at the midpoint of the two 5-year periods prior to the base year.

(3) Multiply the average female populations from step (2) by the age-specific fertility rates pertaining to each 5-year period. The fertility rates reproduce the total fertility rates interpolated for the midpoint of each 5-year period. Sum the resulting births to derive the total number of births during the two 5-year periods prior to the base year.
(4) Separate the total births calculated in step (3) into male and female births using the sex ratio at birth.

(5) Project the births of each sex to the base year using survivorship ratios pertaining to each period for the corresponding ages. This provides a preliminary adjusted population for ages 0 to 4 years and 5 to 9 years.

(6) Smooth the population age 10 years and over using the light smoothing formula developed by Arriaga (see chapter II), which does not change the enumerated population within each 10-year age group. Include as part of the smoothing process the preliminary adjusted population under age 10 years as calculated in step (5).

(7) Repeat the process described above, but this time use the smoothed female population from step (6). Continue the iterations until there is no further change in the figures for the population under age 10 years.

Adjusting only ages under 5 years

This process is similar to the one described above for adjusting the population under age 10 years, but only the estimated population under age 5 years based on births during the 5 years prior to the base year is accepted. The census population for ages 5 to 9 years is taken as enumerated. Hence, the total population over age 5 years is the same as enumerated, although the population at ages 10 years and over is smoothed.

Advantages

This technique produces a base population that, in addition to being smoothed, has a population under age 10 years (or under age 5 years) that is consistent with the mortality and fertility levels for the previous 5 or 10 years.

Limitations

(1) The light and strong smoothing processes should be used with caution. Actual peculiarities of the age structure may be mistakenly smoothed out.

(2) If the estimates of mortality and fertility prior to the base year are not correct, the age structure for the base population will not be correct either. This is a problem related to the mortality and fertility estimates and not a limitation of the technique itself. Nevertheless, the technique should be used with caution when demographic estimates are not robust.

349
Software

The Bureau of the Census has developed two spreadsheets that make the calculations described above. They are BASEPOP and BPSTRNG. BASEPOP adjusts the population under age 10 years and makes a light smoothing of the population in other ages. When it is desired to calculate only the population under age 5 years, the results of BASEPOP are used, but the enumerated population age 5 to 9 years is kept from the census. BPSTRNG follows the same procedures as BASEPOP but makes a strong smoothing of the population over age 10 years. Documentation of the two spreadsheets is presented in volume II.
Appendix VIII-4

Use of the Logistic Function on Demographic Indices

The logistic function can be used for interpolating or extrapolating the trend of demographic indices. There are two alternatives for fitting a logistic function to available data, depending upon whether or not the asymptotic values are known.

If the asymptotes are known or if reasonable asymptotic values can be assumed, the procedure uses logits and the least squares technique based on two or more observed values of the index. If the asymptotes are not known, the procedure requires that three (or a multiple of three) observed data points be available, and that they be equidistant in time. The first alternative is the most frequently used. Both procedures for fitting a logistic are described below.

Fitting a Logistic Function When Asymptotes Are Known or Assumed

Description

This procedure fits a logistic function to values of an index (for example, life expectancies at birth, total fertility rates, or proportions of urban population), given at least two observed values of the index and the two asymptotes of the logistic.

Data Required

1. Two or more values of the index.
2. The lower and upper asymptotes. These can be assumed, as mentioned above. For instance: for life expectancies at birth, 25 years for the lower asymptote for each sex, and 81 years and 87 years, for males and females, respectively, for the upper asymptotes; for total fertility rates, 2 births per woman and eight births per woman, for lower and upper asymptotes, respectively; for percent urban, 0 percent and 90 percent for lower and upper asymptotes, respectively.

Assumptions

The procedure assumes that the trend of the index values follows closely the shape of the logistic function.
Procedure

Assuming that the data to which the logistic function will be fitted follow approximately the shape of a logistic, the logits of the data will follow almost the shape of a straight line because the logits linearize the logistic function. Since the asymptotes are known, a straight line can be fitted to the logits of the data, solving for the two unknowns of the straight line. Under these circumstances, the straight line of the logits is a logistic. The steps are:

1. Apply the following formula to the data:
   \[ Z^t = \ln\left(\frac{U-I^t}{I^t-L}\right) - a + wt \]
   Where:
   - \( Z \) is the logit of the index;
   - \( I^t \) are the indices;
   - \( U \) and \( L \) are the upper and lower asymptotes;
   - \( t \) is the time variable; and
   - \( a \) and \( w \) are the constants to be estimated.

2. To estimate \( a \) and \( w \), apply least squares to the \( Z^t \) values calculated in step (1).

3. With the values of \( a \) and \( w \), calculate the desired values of \( Z^t \), by interpolating or extrapolating.

4. Take the exponential of the \( Z \) values (represented by \( H \)):
   \[ H^t = \exp(Z^t) \]

5. Then, the indices are:
   \[ I^t = \frac{(U + L \cdot H^t)}{(1 + H^t)} \]
   Where the symbols are as above.

Advantages

1. Several demographic indices have a logistic trend within the proper asymptotic values.

2. The logistic function can be used for any of three purposes: to interpolate, to extrapolate, or to determine a trend.
Limitations

Inconsistencies may emerge when the trend of an index is extrapolated independently for each sex.

Software

The Bureau of the Census has developed two spreadsheets to fit a logistic function: EOLGST and LOGISTIC. Either can be used for any index, but the first one is recommended for use with life expectancies at birth. EOLGST fits one logistic to the values of life expectancy at birth for males, and another for females simultaneously in the same spreadsheet. It also calculates the sex differentials of the life expectancies and uses a sex ratio at birth to estimate the life expectancy at birth for both sexes combined. Documentation of these spreadsheets is presented in volume II.

Fitting a Logistic Function When Asymptotes Are Unknown

Description

This procedure fits a logistic function to three (or a multiple of three) observed equidistant values of life expectancies at birth or other indices.

Data Required

Three or a multiple of three equidistant index values.

Assumptions

It is assumed that the trend of the index values follows closely the shape of the logistic function. If it does not, the fitting cannot be made.

Procedure

Estimate the constants of the logistic function by taking the first and second differences of the observed values. In the case of three equidistant observations, and if the logistic \( y \) has the form of

\[
y(i) = \frac{k}{1 + \exp(a + bi)}
\]

Where:

\( k, a, \) and \( b \) are constants; and
i refers to time.

Then follow these steps:

(1) Take the reciprocal (Y) of the observed values:
\[ Y(i) = \frac{1}{y(i)} \]

(2) Take the first and second order differences of the reciprocal values:
\[ D(1) = Y(2) - Y(1) \]
\[ D(2) = Y(3) - Y(2) \]
\[ T = D(2) - D(1) \]

(3) Then, calculate the constants k, a, and b as follows:
\[ k = \frac{T}{T \cdot Y(1) - [D(2)]^2} \]
\[ a = \ln \left\{ \frac{[k \cdot Y(1) - 1]}{k \cdot \exp(b)} \right\} \]
\[ b = \ln \left\{ \frac{d(2)}{D(1)} \right\} \]

Where the symbols are as above.

(4) Once the three constants are determined, use the function to interpolate or extrapolate.

To fit the logistic for more than three observed values, take the reciprocal of each of the values and then group the information into three groups. Follow the same procedure to determine the values of the constants as described above for three data points. Take the first and second differences of grouped reciprocals of the data, and find the constants by algebraic procedures.

**Advantage**

Asymptotes are not required.

**Limitations**

1. The data points must be equidistant in time.

2. The asymptotes implied by fitting the logistic to the demographic indices may not be reasonable from a demographic point of view.
(3) Small variations in the data may have a large impact on the asymptotes and hence on the extrapolated values.

(4) The extrapolation of certain demographic indices may result in inconsistent values. This may occur, for example, when extrapolating life expectancies independently for males and females.

**Software**

The Bureau of the Census has developed a spreadsheet to calculate the logistic according to the procedures described. It is called FITLGSTC, and its documentation is presented volume II.
Appendix VIII-5

Projecting the Trend and Patterns of Mortality

As noted in the text, the projection of mortality requires two principal steps: (a) projecting the general level of mortality, and (b) projecting the pattern of mortality.

To project the general level of mortality, a level for some future date is determined first. Then, mortality levels between the base year and the accepted future level are calculated by interpolation using a logistic function (see appendix VIII-4).

Projecting a future pattern of mortality is more difficult. The pattern must reflect the expected relationships by age and sex: (a) declining mortality from birth to age 10 to 14 years, and continuously increasing mortality thereafter; and (b) in general, female mortality lower than male mortality in each age group. Although mortality may deviate from the expected pattern due to specific epidemics, a surge of violence among specific members of society, or cultural factors affecting the mortality of one sex more than the other, it is usually impossible to predict such conditions for the future.

The projection of the pattern of mortality for each future level of life expectancy at birth may produce inconsistencies, particularly for populations expecting to achieve low levels of mortality in the future. This is especially true when the patterns for males and females are projected independently; in this case, a comparison of the patterns by sex usually reveals the inconsistencies.

Three alternative procedures are recommended for projecting a pattern of mortality to fit the projected levels of life expectancy at birth in developing countries: (a) the direct use of model life tables; (b) the use of the change between successive model life tables; and (c) the use of an interpolation procedure. Among them, the latter procedure usually produces the most acceptable results, provided that the life tables used for interpolation are also acceptable.

Direct Use of Model Life Tables

This is the simplest of the three procedures. Having projected the levels of life expectancy at birth, the pattern of mortality from a model life table is accepted for each level. An inconsistency in the results may occur if life expectancy at birth is considerably higher for males than for females, particularly at low levels of mortality. In such cases, the sex differentials of infant mortality may be too large. (See examples in table A-VIII-5.1.)

The United Nations has developed the program MATCH in the MORTPAK package, which can be used for generating model life tables.
Table A-VIII-5.1. Comparison of Several Procedures for Predicting the Pattern of Mortality
(Base Data: Mexican Life Table for 1950.)

<table>
<thead>
<tr>
<th>Year</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Life expectancy at birth</td>
<td>Projected infant mortality rates with indicated procedures</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1950</td>
<td>46.16</td>
<td>49.00</td>
</tr>
<tr>
<td>1980</td>
<td>64.70</td>
<td>70.34</td>
</tr>
<tr>
<td>2000</td>
<td>72.86</td>
<td>79.11</td>
</tr>
</tbody>
</table>

Notes:
B: Coale-Demeny model life tables region West.
C: United Nations logit procedure for adjusting empirical pattern to another level of life expectancy at birth.
D: Interpolation procedure between base life table and "limit" life table.
E: The relative change of the U. N. model general pattern (A) is directly applied to the base life table.
F: Same as (E), but using the relative change of the Coale-Demeny life table region West (B).

Calculations for A, B, C, were made using MATCH from the United Nations computer package MORTPAK.
Calculations for D were made using U. S. Bureau of the Census spreadsheets ITERLTF and ITERLTM.
Use of the Change Between Successive Model Life Tables

As discussed in the text, this procedure starts by generating model life tables for the projected levels of life expectancy at birth. Then, the observed changes in mortality between successive life tables are applied to an empirical life table pertaining to the population being studied. The purpose of the procedure is to maintain the country's own pattern of mortality. This may be accomplished in one of two ways: (a) by using the relative change in the central mortality rates of the models and applying it to the country's pattern; or (b) by using a procedure developed by the United Nations and presented in the computer program MATCH of the MORTPAK package. (See examples in table A-VIII-5.1.)

Use of an Interpolation Procedure

In this procedure, interpolations are made between the base life table and a hypothetical life table with very low mortality. The interpolation is performed with logits or logarithms of the probabilities of dying or central death rates of the life tables. The procedure is iterative and produces an interpolated life table with the desired level of life expectancy at birth. If both life tables are acceptable, the interpolated values between them should also be acceptable. (See examples in table A-VIII-5.1.) The Bureau of the Census has developed two spreadsheets, one for each sex, for interpolating life tables. The spreadsheets are called INTPLTM and INTPLTF; see below.

Examples

Examples are presented below for the three alternative procedures for projecting the pattern of mortality. They pertain to Mexico and Nepal. For both countries, life tables for a given year are available. For Mexico, a life table for 1950 is taken as a basis for the comparison, while for Nepal the base life table pertains to 1975. Mexico has a rather large sex differential in mortality, while Nepal has a higher life expectancy for males than for females. Alternative cases A through F are presented for each country, some of them illustrating inconsistencies that may occur in the sex differentials of infant and child mortality under the country's circumstances.

Mexico. Empirical life tables are available for several years. Based on a life table for 1950, patterns of mortality are estimated for 1980 and 2000, based on the observed level of life expectancy at birth in 1980 (from an empirical life table), and a projected level of life expectancy for the year 2000. Only infant and child mortality for each sex is compared among the alternatives. See table A-VIII-5.1.

In the Mexico examples, the estimated values for each sex do not show great inconsistencies, except for procedures (B) and (F). In these two procedures, the sex differential becomes too large for infant and child mortality at levels of life expectancy at birth of 72.86 years and 79.11 years for males and females, respectively. Case (B) represents Coale-Demeny West region model life tables, and case (F) represents the application of the
change from model life tables to empirical values. These two estimation procedures are closely related.

Nepal. An empirical life table is available for 1975, showing a higher life expectancy for males than for females. Projected life expectancies at birth for 2000 and 2050 are used for projecting the pattern of mortality. See table A-VIII-5.2.

In this case, the same inconsistencies as for Mexico appear at the highest levels of life expectancy. In addition, infant mortality corresponding to the male and female levels of life expectancy at birth of 57 years present some questionable estimates. For example, infant mortality is higher for females than for males in procedures (B) and (C). Although this may not be wrong for Nepal's population, it seldom occurs at this level of life expectancy at birth.

Spreadsheets INTPLTM and INTPLTF

Description

The purpose of these spreadsheets is to interpolate a male and female life table, respectively, between the values of two given sets of "pivotal" life tables. The interpolation is based on the probabilities of dying, or $nq_x$ values, for each sex. The procedure iterates until the interpolated life table has the exact desired life expectancy at birth. Usually, the two pivotal life tables are: (a) an empirical life table for the base year of the projection, and (b) an ultimate (or upper limit) life table. The spreadsheet developed at the Bureau of the Census provides male and female ultimate life tables based on empirical mortality information by sex, age, and causes of death for developed countries. It allows the user to change these ultimate life tables. The user must provide the high mortality life table accepted for the population being analyzed.

Data Required

The following information from a life table for each sex (usually an empirical one for the base year):

(1) The probabilities of dying between exact ages, by 5-year age groups for ages 5 years and above, and for ages under 1 year and 1 to 4 years.

(2) Life expectancy at birth (preferably with two or three decimal places).

(3) Separation factors for ages under 1 year and 1 to 4 years.

(4) The central death rate of the open-ended age group.
Table A-VIII-5.2. Comparison of Several Procedures for Predicting the Pattern of Mortality
(Base Data: Nepal Life Table for 1975.)

<table>
<thead>
<tr>
<th>Year</th>
<th>Life expectancy at birth</th>
<th>Projected infant mortality rates with indicated procedures</th>
<th>Projected central death rates for the age group 1 to 4 years, with indicated procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>1975</td>
<td>44.71</td>
<td>41.76</td>
<td>144.50</td>
</tr>
<tr>
<td>2000</td>
<td>57.00</td>
<td>57.00</td>
<td>84.75</td>
</tr>
<tr>
<td>2050</td>
<td>74.29</td>
<td>80.00</td>
<td>21.22</td>
</tr>
</tbody>
</table>

Notes:
B: Coale-Demeny model life tables region West.
C: United Nations logit procedure for adjusting empirical pattern to another level of life expectancy at birth.
D: Interpolation procedure between base life table and "limit" life table.
E: The relative change of the U. N. model general pattern (A) is directly applied to the base life table.
F: Same as (E), but using the relative change of the Coale-Demeny life table region West (B).

Calculations for A, B, C, were made using MATCH from the United Nations computer package MORTPAK.
Calculations for D were made using U. S. Bureau of the Census spreadsheets ITERLTF and ITERLTM.
### Table A-VIII-5.3. Ultimate Life Table for Males and Females

<table>
<thead>
<tr>
<th>Age</th>
<th>Width</th>
<th>nMx</th>
<th>nax</th>
<th>nqx</th>
<th>lx</th>
<th>ndx</th>
<th>5Px</th>
<th>Tx</th>
<th>ex</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>.00357</td>
<td>.050</td>
<td>.00356</td>
<td>100,000</td>
<td>356</td>
<td>.99627</td>
<td>8,155,531</td>
<td>81.56</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>.00011</td>
<td>1.640</td>
<td>.00044</td>
<td>99,644</td>
<td>44</td>
<td>.99951</td>
<td>8,055,869</td>
<td>80.85</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>.00009</td>
<td>2.500</td>
<td>.00065</td>
<td>99,600</td>
<td>45</td>
<td>.99960</td>
<td>7,657,397</td>
<td>76.88</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>.00007</td>
<td>2.500</td>
<td>.00335</td>
<td>99,555</td>
<td>35</td>
<td>.99893</td>
<td>7,159,508</td>
<td>71.91</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>.00036</td>
<td>2.500</td>
<td>.01800</td>
<td>99,521</td>
<td>179</td>
<td>.99778</td>
<td>6,661,818</td>
<td>66.94</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>.00033</td>
<td>2.500</td>
<td>.02650</td>
<td>99,342</td>
<td>263</td>
<td>.99760</td>
<td>6,164,663</td>
<td>62.06</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>.00024</td>
<td>2.500</td>
<td>.02150</td>
<td>99,079</td>
<td>213</td>
<td>.99780</td>
<td>5,668,612</td>
<td>57.21</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>.00045</td>
<td>2.500</td>
<td>.00225</td>
<td>98,866</td>
<td>222</td>
<td>.99706</td>
<td>5,173,751</td>
<td>52.33</td>
</tr>
<tr>
<td>35</td>
<td>5</td>
<td>.00073</td>
<td>2.500</td>
<td>.00364</td>
<td>98,644</td>
<td>359</td>
<td>.99544</td>
<td>4,679,978</td>
<td>47.44</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>.00110</td>
<td>2.500</td>
<td>.00548</td>
<td>98,284</td>
<td>539</td>
<td>.99241</td>
<td>4,187,658</td>
<td>42.61</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
<td>.00195</td>
<td>2.500</td>
<td>.00970</td>
<td>97,745</td>
<td>948</td>
<td>.98659</td>
<td>3,697,584</td>
<td>37.83</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>.00330</td>
<td>2.500</td>
<td>.01636</td>
<td>96,797</td>
<td>1,584</td>
<td>.97779</td>
<td>3,211,229</td>
<td>33.17</td>
</tr>
<tr>
<td>55</td>
<td>5</td>
<td>.00371</td>
<td>2.500</td>
<td>.02815</td>
<td>95,213</td>
<td>2,680</td>
<td>.96477</td>
<td>2,731,206</td>
<td>28.69</td>
</tr>
<tr>
<td>60</td>
<td>5</td>
<td>.00869</td>
<td>2.500</td>
<td>.04253</td>
<td>92,533</td>
<td>3,935</td>
<td>.94449</td>
<td>2,261,842</td>
<td>24.44</td>
</tr>
<tr>
<td>65</td>
<td>5</td>
<td>.01431</td>
<td>2.500</td>
<td>.06908</td>
<td>88,598</td>
<td>6,120</td>
<td>.90751</td>
<td>1,809,017</td>
<td>20.42</td>
</tr>
<tr>
<td>70</td>
<td>5</td>
<td>.02500</td>
<td>2.500</td>
<td>.11765</td>
<td>82,477</td>
<td>9,703</td>
<td>.84702</td>
<td>1,381,329</td>
<td>16.75</td>
</tr>
<tr>
<td>75</td>
<td>5</td>
<td>.04273</td>
<td>2.500</td>
<td>.19303</td>
<td>72,774</td>
<td>14,048</td>
<td>.76621</td>
<td>993,200</td>
<td>13.65</td>
</tr>
<tr>
<td>80</td>
<td>5</td>
<td>.06628</td>
<td>2.500</td>
<td>.28429</td>
<td>58,727</td>
<td>16,696</td>
<td>.62090</td>
<td>664,449</td>
<td>11.31</td>
</tr>
<tr>
<td>85+</td>
<td></td>
<td>.10188</td>
<td>9.815</td>
<td>1.00000</td>
<td>42,031</td>
<td>42,031</td>
<td>.412,554</td>
<td></td>
<td>9.82</td>
</tr>
</tbody>
</table>

- nMx = Age-specific central death rate.
- nax = Average person-years lived by those who die between ages x and x+n.
- nqx = Probability of dying between exact ages x and x+n (age-specific mortality rate).
- lx = Number of survivors at age x.
- ndx = Number of deaths occurring between ages x and x+n.
- nLx = Number of person-years lived between ages x and x+n.
- 5Px = Survival ratio for persons aged x to x+5 surviving 5 years to ages x+5 to x+10 = 5Lx+5/5Lx (first 5Px = 5L0/5L0, second 5Px = 5L5/5L0, last 5Px = Tx+5/Tx).
- Tx = Number of person-years lived after age x.
- ex = Life expectancy at age x.
### Table A-VIII-5.3. Ultimate Life Table for Males and Females. Continued...

<table>
<thead>
<tr>
<th>Age, x</th>
<th>Width, n</th>
<th>nMx</th>
<th>nax</th>
<th>nqx</th>
<th>lx</th>
<th>ndx</th>
<th>nlx</th>
<th>5Px</th>
<th>Tx</th>
<th>ex</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.00301</td>
<td>0.060</td>
<td>0.00300</td>
<td>100</td>
<td>99,718</td>
<td>0.99690</td>
<td>8,723,342</td>
<td>87.23</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0.000007</td>
<td>1.520</td>
<td>0.0028</td>
<td>99,700</td>
<td>28</td>
<td>398,730</td>
<td>0.9972</td>
<td>8,623,624</td>
<td>86.50</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.00400</td>
<td>2.500</td>
<td>0.000020</td>
<td>99,672</td>
<td>20</td>
<td>496,310</td>
<td>0.99085</td>
<td>8,224,894</td>
<td>82.52</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>0.000020</td>
<td>2.500</td>
<td>0.00010</td>
<td>99,652</td>
<td>10</td>
<td>496,235</td>
<td>0.9979</td>
<td>7,726,584</td>
<td>77.54</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>0.00006</td>
<td>2.500</td>
<td>0.00031</td>
<td>99,642</td>
<td>31</td>
<td>496,132</td>
<td>0.99565</td>
<td>7,228,349</td>
<td>72.54</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>0.00011</td>
<td>2.500</td>
<td>0.00057</td>
<td>99,611</td>
<td>57</td>
<td>497,912</td>
<td>0.99355</td>
<td>6,730,216</td>
<td>67.57</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>0.00015</td>
<td>2.500</td>
<td>0.00073</td>
<td>99,554</td>
<td>73</td>
<td>497,509</td>
<td>0.99145</td>
<td>6,232,304</td>
<td>62.60</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>0.00020</td>
<td>2.500</td>
<td>0.00099</td>
<td>99,481</td>
<td>99</td>
<td>497,161</td>
<td>0.98665</td>
<td>5,734,715</td>
<td>57.65</td>
</tr>
<tr>
<td>35</td>
<td>5</td>
<td>0.00034</td>
<td>2.500</td>
<td>0.00169</td>
<td>99,383</td>
<td>168</td>
<td>496,493</td>
<td>0.97625</td>
<td>5,237,554</td>
<td>52.70</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>0.00062</td>
<td>2.500</td>
<td>0.00308</td>
<td>99,215</td>
<td>305</td>
<td>495,310</td>
<td>0.95755</td>
<td>4,741,061</td>
<td>47.79</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
<td>0.00109</td>
<td>2.500</td>
<td>0.00543</td>
<td>98,909</td>
<td>537</td>
<td>493,205</td>
<td>0.92665</td>
<td>4,245,751</td>
<td>42.93</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.00186</td>
<td>2.500</td>
<td>0.00926</td>
<td>98,373</td>
<td>911</td>
<td>489,585</td>
<td>0.87965</td>
<td>3,752,546</td>
<td>38.15</td>
</tr>
<tr>
<td>55</td>
<td>5</td>
<td>0.00299</td>
<td>2.500</td>
<td>0.01685</td>
<td>97,461</td>
<td>1,448</td>
<td>483,608</td>
<td>0.80886</td>
<td>3,262,962</td>
<td>33.48</td>
</tr>
<tr>
<td>60</td>
<td>5</td>
<td>0.00475</td>
<td>2.500</td>
<td>0.02349</td>
<td>96,014</td>
<td>2,255</td>
<td>474,431</td>
<td>0.69098</td>
<td>2,779,274</td>
<td>28.95</td>
</tr>
<tr>
<td>65</td>
<td>5</td>
<td>0.00786</td>
<td>2.500</td>
<td>0.03854</td>
<td>93,759</td>
<td>3,613</td>
<td>459,762</td>
<td>0.48444</td>
<td>2,304,843</td>
<td>24.58</td>
</tr>
<tr>
<td>70</td>
<td>5</td>
<td>0.01346</td>
<td>2.500</td>
<td>0.06511</td>
<td>90,146</td>
<td>5,869</td>
<td>436,056</td>
<td>0.11125</td>
<td>1,845,081</td>
<td>20.47</td>
</tr>
<tr>
<td>75</td>
<td>5</td>
<td>0.02392</td>
<td>2.500</td>
<td>0.11285</td>
<td>84,276</td>
<td>9,511</td>
<td>397,605</td>
<td>0.08821</td>
<td>1,409,025</td>
<td>16.72</td>
</tr>
<tr>
<td>80</td>
<td>5</td>
<td>0.04338</td>
<td>2.500</td>
<td>0.19568</td>
<td>74,766</td>
<td>14,630</td>
<td>337,253</td>
<td>0.06655</td>
<td>1,011,420</td>
<td>13.53</td>
</tr>
<tr>
<td>85</td>
<td>+</td>
<td>0.08920</td>
<td>11.211</td>
<td>1.00000</td>
<td>60,136</td>
<td>60,136</td>
<td>674,167</td>
<td>0.0674</td>
<td>1,674,167</td>
<td>11.21</td>
</tr>
</tbody>
</table>

### Notes:
- **nMx** = Age-specific central death rate.
- **nax** = Average person-years lived by those who die between ages x and x+n.
- **nqx** = Probability of dying between exact ages x and x+n (age-specific mortality rate).
- **lx** = Number of survivors at age x.
- **ndx** = Number of deaths occurring between ages x and x+n.
- **nlx** = Number of person-years lived between ages x and x+n.
- **5Px** = Survival ratio for persons aged x to x+5 surviving 5 years to ages x+5 to x+10 = 5Lx+5/5Lx (first 5Px = 5L0/5L0, second 5Px = 5L5/5L0, last 5Px = Tx+5/Tx).
- **Tx** = Number of person-years lived after age x.
- **ex** = Life expectancy at age x.

Separation factors: Empirical.

### Remarks:
- For ages 15+ based on a moving average of the logs: smoothed log(5Mx) = 1/3 [log(5Mx-5)+log(5Mx)+log(5Mx+5)].

---

363
**Assumptions**

The interpolation procedure assumes that logarithms of the age-specific mortality rates change linearly between the logarithms of the corresponding rates of the two pivotal life tables in relation to the changes in life expectancies at birth.

**Procedure**

The interpolated life table is obtained by the following steps:

1. The logarithms of the probabilities of dying in the two pivotal life tables are linearly interpolated in relation to the life expectancies of the pivotal life tables and the life expectancy desired for the interpolated life table.

2. A new life table is constructed with the interpolated probabilities of dying.

3. The life expectancy of the life table constructed in step (2) is compared with the desired life expectancy, and the relative difference is used to proportionally adjust the probabilities of dying.

4. Using the adjusted probabilities of dying from step (3), another life table is constructed, and its life expectancy at birth is compared with the desired life expectancy.

5. In an iterative process, steps (3) and (4) are repeated until the constructed life table has the desired life expectancy. Usually no more than 20 iterations are required.

**Advantages**

1. The ultimate, or upper limit, life table can be easily changed, if desired, according to the needs of the researcher.

2. Because the two pivotal life tables are selected by the user, the patterns of the interpolated life tables should be consistent and practically free of unexpected results.
Limitations

This procedure is recommended only for interpolating between the base and ultimate life tables. Extrapolated life tables outside of this range may not be acceptable.

Software

The documentation of these two spreadsheets (INTPLTM and INTPLTF) is presented in volume II.
Appendix VIII-6

Projecting the Trend of Fertility

As in the case of mortality, there are two steps for projecting the trend of fertility: (a) to determine a trend based on past levels of total fertility rates and extrapolating them; and (b) to estimate a future level of fertility and interpolate between it and the level available for the beginning of the population projection.

For the first step, there are two programs based on the logistic function: TFRLGST and LOGISTIC. The calculations made by both programs are the same. However, the instructions for data input in TFRLGST refer specifically to fertility (see volume II), while the documentation of the LOGISTIC program is generalized for the use of any index (see appendix VIII-4 and volume II).

For the second step, the same two programs can be used for the interpolation process, TFRLGST and LOGISTIC (see documentation in volume II). Another possibility is to use the sine function as presented in spreadsheet TFRSINE. The sine function starts the interpolation process with a small change of the index, which could be an advantage over the logistic function in cases where an expected decline in fertility has not yet begun. Documentation of the TFRSINE spreadsheet is presented in volume II.
Appendix VIII-7

Projecting the Pattern of Age-Specific Fertility Rates

Introduction

Projecting a pattern of age-specific fertility rates has a certain similarity to projecting a pattern of mortality. Age-specific fertility rates normally increase up to a particular age, usually between 20 and 30 years of age. After achieving the maximum value, they decline to zero around age 50 years.

Description

The pattern of fertility is interpolated between empirical age-specific fertility rates pertaining to particular levels of total fertility rates. Given a desired level of total fertility rate, the procedure interpolates between two sets of age-specific fertility rates and provides a new set that corresponds to the desired total fertility rate.

Data Required

The total fertility rate for which a set of age-specific fertility rates is needed.

Assumptions

The pattern of fertility pertaining to the total fertility rate is similar to the patterns used for interpolation purposes.

Procedure

1. To develop this procedure, various age patterns of fertility were derived based on empirical data. Age-specific fertility rates from developing countries were grouped according to their levels of total fertility rates, namely 1, 2, 3, 4, 5, 6, 7, and 8 births per woman. For each of these groups, an average of the age-specific fertility rates was taken. Patterns for total fertility rates outside this range cannot be accommodated.

2. In this technique, interpolation of age-specific fertility rates is performed in relation to levels of total fertility rates. The "pivotal" rates used in the interpolation process are the two sets pertaining to the levels of total fertility rates (integers) that are closest to the desired level. For example, if the desired set of age-specific fertility rates is to correspond to a total fertility rate of 5.23 births per woman, the interpolation is made...
between the average fertility rates for the levels of total fertility rates of 6 and 5 births per woman.

(3) Once the age-specific fertility rates have been interpolated, they are proportionally adjusted to obtain the desired level of total fertility rate.

Advantages

The technique requires only a total fertility rate.

Limitations

There is no control for the age of maximum fertility, and hence the derived pattern of fertility may be different from the actual pattern of the population being studied. However, the patterns of fertility used in the computer program can be changed at will.

Software

The Bureau of the Census has developed a spreadsheet that performs this procedure. It is called ASFRPATT, and its documentation is presented in volume II.
Appendix VIII-8

The Impact of Errors in the Base Population of a Projection

This appendix illustrates the impact that errors in the base population may produce in a population projection. For this purpose, two population projections were made. Both have the same levels of mortality and fertility, and they are not exposed to migration. The only difference between the two projections is the age structure of the base populations:

(1) The first projection started with an age structure for the base year that was affected by age misreporting and underenumeration of the population under age 10 years. This represents the unsmoothed population.

(2) The second projection started with a base population which had been smoothed. The smoothing process was applied to the population 10 years of age and over to correct for age misreporting. In addition, the population under 10 years of age was estimated taking into account the levels of fertility and mortality during the 10-year period prior to the census date. This represents the smoothed population.

The unsmoothed age distribution in this example resembles that of a developing country population. For simplicity, the pattern of mortality was assumed to be the same as the model life tables of Coale-Demeny, region South. The age-specific fertility rates were derived from the pattern of fertility in the spreadsheet ASFRPATT presented in figure VIII-8 and appendix VIII-7. The levels of mortality and fertility, measured by the life expectancies at birth and total fertility rates, respectively, are given below:

<table>
<thead>
<tr>
<th>Year</th>
<th>Life expectancy at birth</th>
<th>Total fertility rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>1982</td>
<td>49.94</td>
<td>52.82</td>
</tr>
<tr>
<td>1985</td>
<td>51.19</td>
<td>54.40</td>
</tr>
<tr>
<td>1990</td>
<td>53.26</td>
<td>57.02</td>
</tr>
<tr>
<td>1995</td>
<td>55.33</td>
<td>59.62</td>
</tr>
<tr>
<td>2000</td>
<td>57.35</td>
<td>62.16</td>
</tr>
<tr>
<td>2005</td>
<td>59.31</td>
<td>64.61</td>
</tr>
<tr>
<td>2010</td>
<td>61.20</td>
<td>66.93</td>
</tr>
<tr>
<td>2015</td>
<td>62.98</td>
<td>69.11</td>
</tr>
<tr>
<td>2020</td>
<td>64.66</td>
<td>71.13</td>
</tr>
</tbody>
</table>

The results of the population projections for selected years and the differences between the two projections are presented in tables A-VIII-8.1, A-VIII-8.2, A-VIII-8.3, and A-VIII-8.4 of this appendix. Age pyramids of the projected populations for the same selected years were presented in the text (figure VIII-9).
Table A-VIII-8.1. Unsmoothed and Smoothed Base Population for 1982

<table>
<thead>
<tr>
<th>Age group</th>
<th>Unsmoothed population</th>
<th>Smoothed population</th>
<th>Differences (unsmoothed minus smoothed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Total</td>
<td>1,247,547</td>
<td>1,317,878</td>
<td>1,300,093</td>
</tr>
<tr>
<td>0-4</td>
<td>202,845</td>
<td>199,398</td>
<td>261,913</td>
</tr>
<tr>
<td>5-9</td>
<td>208,804</td>
<td>200,865</td>
<td>202,822</td>
</tr>
<tr>
<td>10-14</td>
<td>190,364</td>
<td>134,476</td>
<td>162,471</td>
</tr>
<tr>
<td>15-19</td>
<td>127,120</td>
<td>118,657</td>
<td>131,193</td>
</tr>
<tr>
<td>20-24</td>
<td>78,564</td>
<td>162,972</td>
<td>104,713</td>
</tr>
<tr>
<td>25-29</td>
<td>78,890</td>
<td>129,269</td>
<td>86,545</td>
</tr>
<tr>
<td>30-34</td>
<td>88,497</td>
<td>83,849</td>
<td>72,673</td>
</tr>
<tr>
<td>35-39</td>
<td>63,302</td>
<td>60,390</td>
<td>60,776</td>
</tr>
<tr>
<td>40-44</td>
<td>53,038</td>
<td>59,143</td>
<td>50,408</td>
</tr>
<tr>
<td>45-49</td>
<td>33,075</td>
<td>43,517</td>
<td>40,567</td>
</tr>
<tr>
<td>50-54</td>
<td>33,322</td>
<td>35,236</td>
<td>29,820</td>
</tr>
<tr>
<td>55-59</td>
<td>15,664</td>
<td>14,371</td>
<td>23,346</td>
</tr>
<tr>
<td>60-64</td>
<td>23,802</td>
<td>24,852</td>
<td>18,592</td>
</tr>
<tr>
<td>65-69</td>
<td>10,999</td>
<td>13,650</td>
<td>14,633</td>
</tr>
<tr>
<td>70-74</td>
<td>12,283</td>
<td>11,595</td>
<td>11,188</td>
</tr>
<tr>
<td>75-79</td>
<td>7,160</td>
<td>6,851</td>
<td>8,255</td>
</tr>
<tr>
<td>80+</td>
<td>20,718</td>
<td>18,777</td>
<td>20,718</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of the Census.
<table>
<thead>
<tr>
<th>Age group</th>
<th>Unsmoothed Male</th>
<th>Unsmoothed Female</th>
<th>Smoothed Male</th>
<th>Smoothed Female</th>
<th>Differences (unsmoothed minus smoothed) Male</th>
<th>Differences (unsmoothed minus smoothed) Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1,844,098</td>
<td>1,916,238</td>
<td>1,866,386</td>
<td>1,952,641</td>
<td>-22,288</td>
<td>-36,403</td>
</tr>
<tr>
<td>0-4</td>
<td>382,723</td>
<td>377,994</td>
<td>376,726</td>
<td>372,055</td>
<td>5,997</td>
<td>5,939</td>
</tr>
<tr>
<td>5-9</td>
<td>326,463</td>
<td>322,983</td>
<td>309,400</td>
<td>305,972</td>
<td>17,263</td>
<td>17,011</td>
</tr>
<tr>
<td>10-14</td>
<td>189,033</td>
<td>186,475</td>
<td>243,398</td>
<td>244,289</td>
<td>-54,365</td>
<td>-57,814</td>
</tr>
<tr>
<td>15-19</td>
<td>203,606</td>
<td>196,056</td>
<td>197,238</td>
<td>198,120</td>
<td>6,368</td>
<td>-2,064</td>
</tr>
<tr>
<td>20-24</td>
<td>184,053</td>
<td>130,976</td>
<td>157,763</td>
<td>152,229</td>
<td>27,000</td>
<td>-21,253</td>
</tr>
<tr>
<td>25-29</td>
<td>122,207</td>
<td>114,675</td>
<td>126,100</td>
<td>128,506</td>
<td>-3,893</td>
<td>-13,831</td>
</tr>
<tr>
<td>30-34</td>
<td>75,088</td>
<td>156,722</td>
<td>100,077</td>
<td>121,795</td>
<td>-24,989</td>
<td>34,927</td>
</tr>
<tr>
<td>35-39</td>
<td>75,045</td>
<td>123,891</td>
<td>82,342</td>
<td>105,663</td>
<td>-7,297</td>
<td>18,228</td>
</tr>
<tr>
<td>40-44</td>
<td>83,456</td>
<td>79,945</td>
<td>68,537</td>
<td>86,908</td>
<td>14,919</td>
<td>-6,963</td>
</tr>
<tr>
<td>45-49</td>
<td>38,806</td>
<td>57,142</td>
<td>56,450</td>
<td>69,777</td>
<td>2,356</td>
<td>-12,635</td>
</tr>
<tr>
<td>50-54</td>
<td>48,095</td>
<td>55,251</td>
<td>45,698</td>
<td>50,655</td>
<td>2,397</td>
<td>4,596</td>
</tr>
<tr>
<td>55-59</td>
<td>28,073</td>
<td>39,707</td>
<td>35,414</td>
<td>38,420</td>
<td>-6,541</td>
<td>1,207</td>
</tr>
<tr>
<td>60-64</td>
<td>27,383</td>
<td>30,797</td>
<td>24,507</td>
<td>28,389</td>
<td>2,876</td>
<td>2,408</td>
</tr>
<tr>
<td>65-69</td>
<td>11,675</td>
<td>11,569</td>
<td>17,391</td>
<td>20,504</td>
<td>-5,716</td>
<td>-8,935</td>
</tr>
<tr>
<td>70-74</td>
<td>15,010</td>
<td>17,157</td>
<td>11,756</td>
<td>13,766</td>
<td>3,254</td>
<td>3,391</td>
</tr>
<tr>
<td>75-79</td>
<td>4,881</td>
<td>7,376</td>
<td>6,978</td>
<td>8,122</td>
<td>-2,097</td>
<td>-746</td>
</tr>
<tr>
<td>80+</td>
<td>6,901</td>
<td>7,522</td>
<td>6,811</td>
<td>7,471</td>
<td>90</td>
<td>51</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of the Census.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Unsmoothed base population</th>
<th>Smoothed base population</th>
<th>Differences (unsmoothed minus smoothed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Total</td>
<td>2,864,577</td>
<td>2,946,085</td>
<td>2,939,405</td>
</tr>
<tr>
<td>0-4</td>
<td>501,289</td>
<td>495,867</td>
<td>535,349</td>
</tr>
<tr>
<td>5-9</td>
<td>431,662</td>
<td>429,506</td>
<td>451,430</td>
</tr>
<tr>
<td>10-14</td>
<td>389,006</td>
<td>387,177</td>
<td>389,351</td>
</tr>
<tr>
<td>15-19</td>
<td>342,600</td>
<td>340,802</td>
<td>333,036</td>
</tr>
<tr>
<td>20-24</td>
<td>252,930</td>
<td>249,777</td>
<td>275,443</td>
</tr>
<tr>
<td>25-29</td>
<td>196,375</td>
<td>199,302</td>
<td>216,784</td>
</tr>
<tr>
<td>30-34</td>
<td>192,169</td>
<td>166,219</td>
<td>172,931</td>
</tr>
<tr>
<td>35-39</td>
<td>154,852</td>
<td>111,152</td>
<td>137,604</td>
</tr>
<tr>
<td>40-44</td>
<td>93,350</td>
<td>126,285</td>
<td>108,544</td>
</tr>
<tr>
<td>45-49</td>
<td>65,681</td>
<td>143,992</td>
<td>85,768</td>
</tr>
<tr>
<td>50-54</td>
<td>72,895</td>
<td>98,659</td>
<td>69,376</td>
</tr>
<tr>
<td>55-59</td>
<td>65,809</td>
<td>61,962</td>
<td>55,334</td>
</tr>
<tr>
<td>60-64</td>
<td>43,788</td>
<td>49,106</td>
<td>42,511</td>
</tr>
<tr>
<td>65-69</td>
<td>29,692</td>
<td>40,803</td>
<td>30,761</td>
</tr>
<tr>
<td>70-74</td>
<td>17,221</td>
<td>25,673</td>
<td>19,279</td>
</tr>
<tr>
<td>75-79</td>
<td>9,971</td>
<td>12,429</td>
<td>9,859</td>
</tr>
<tr>
<td>80+</td>
<td>5,287</td>
<td>7,374</td>
<td>6,045</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of the Census.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Unsmoothed base population</th>
<th>Smoothed base population</th>
<th>Differences (unsmoothed minus smoothed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Total</td>
<td>4,637,053</td>
<td>4,652,431</td>
<td>4,685,684</td>
</tr>
<tr>
<td>0-4</td>
<td>703,917</td>
<td>697,319</td>
<td>723,287</td>
</tr>
<tr>
<td>5-9</td>
<td>638,207</td>
<td>635,748</td>
<td>658,522</td>
</tr>
<tr>
<td>10-14</td>
<td>652,532</td>
<td>561,242</td>
<td>590,370</td>
</tr>
<tr>
<td>15-19</td>
<td>482,163</td>
<td>482,007</td>
<td>514,954</td>
</tr>
<tr>
<td>20-24</td>
<td>423,391</td>
<td>424,375</td>
<td>442,787</td>
</tr>
<tr>
<td>25-29</td>
<td>379,932</td>
<td>381,593</td>
<td>380,273</td>
</tr>
<tr>
<td>30-34</td>
<td>332,639</td>
<td>334,475</td>
<td>323,164</td>
</tr>
<tr>
<td>35-39</td>
<td>244,012</td>
<td>244,212</td>
<td>265,669</td>
</tr>
<tr>
<td>40-44</td>
<td>187,874</td>
<td>193,851</td>
<td>207,447</td>
</tr>
<tr>
<td>45-49</td>
<td>181,266</td>
<td>160,527</td>
<td>163,163</td>
</tr>
<tr>
<td>50-54</td>
<td>142,725</td>
<td>106,024</td>
<td>126,762</td>
</tr>
<tr>
<td>55-59</td>
<td>82,820</td>
<td>117,987</td>
<td>96,191</td>
</tr>
<tr>
<td>60-64</td>
<td>54,715</td>
<td>130,360</td>
<td>71,541</td>
</tr>
<tr>
<td>65-69</td>
<td>55,114</td>
<td>84,175</td>
<td>52,640</td>
</tr>
<tr>
<td>70-74</td>
<td>42,752</td>
<td>47,015</td>
<td>35,809</td>
</tr>
<tr>
<td>75-79</td>
<td>21,270</td>
<td>29,395</td>
<td>20,716</td>
</tr>
<tr>
<td>80+</td>
<td>11,924</td>
<td>22,126</td>
<td>12,389</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of the Census.
REFERENCES


Chapter IX
THE WRITTEN REPORT

Introduction

A demographic analysis of the information provided by censuses and/or surveys should be made as soon as possible after the data become available. One of the purposes of this manual is to facilitate making such an analysis in a short period of time, once the information is available. The use of the computer programs presented in this manual will help to accelerate the calculation process. But still, the researcher must produce a report about the country's demographic situation and population characteristics so the analysis will be available to others. This chapter presents some guidance on preparing a demographic report.

The report's style will depend primarily on the purpose of the report and the researcher's own preferences. Insofar as possible, the report should keep technical and nontechnical aspects separated. The main text should be nontechnical, so non-demographers can easily understand it. Text presentation of population characteristics should be supported by graphics and small tables presenting the most significant findings. Technical discussion and methodologies used, as well as detailed data tables, should be presented in the appendices.

Nontechnical Report

The report should be clear and concise. It should start with a summary of the main findings considered in the study (mortality, fertility, etc). It should include the impact of the demographic characteristics on population growth and age structure, with references to social aspects such as the potential number of students, size of the labor force, and other population issues. The presentation of this nontechnical section should include graphics, so the levels and trends of the estimated variables can be easily grasped. This section should not include too many numbers nor large tables; only the figures required to present a general overview should be presented.

A short section could present the degree of reliability of the information on population characteristics presented in the report. For instance, one paragraph could discuss possible errors in the estimates because they were based on data that may not have been collected properly, because they pertain to small samples, or because they were calculated indirectly. A full discussion of those problems should be included in an appendix.

Another section should compare the estimated demographic characteristics with similar estimates for other dates during the past. This section may include an analysis of trends and the meaning and consequences of such trends. It may include a comparison of the country's levels with those of other countries in the same region or other areas of the world.
Finally, if the country has a policy related to population issues, a comparative analysis of the estimated levels with those targeted in the policy should be made.

Technical Report

The technical sections of the report (covering population age and sex structure, mortality, etc.) may each follow a similar format. General guidelines are given here for such sections. Guidelines for the section on population projections are given separately because that topic is unique.

General technical presentation

This section should include discussion of topics such as population age structure, mortality, fertility, migration, urbanization, and so forth. Each of them may follow the format suggested below.

Introduction. The introduction should discuss the previous and current estimates of the particular variable (or quality of the age structure from different censuses) including graphs of historical trends. It should follow with a brief evaluation of earlier estimates, the quality of the information used, and/or the techniques or procedures used. Finally, it should mention the data and techniques used for current estimates.

Data availability. This section describes the information used for obtaining new estimates. Reported data used in the estimation process should be given in the appendices for use by other researchers.

Reliability of the data. This section provides a technical discussion of the evaluation of the new and previous data revealing possible errors in collecting the information or in the geographical area representation of the data. If data pertain to samples, standard errors of data could be given.

Methods applied. This section should mention the techniques that could have been applied to the available information and those that actually were applied during the estimation process. It should include:

(1) A brief discussion about why a particular technique was selected.
(2) Possible data errors and their impact on the estimates.
(3) Biases that estimates may have because the data used did not match the assumptions required by the techniques.
(4) An explanation of why a particular estimate was selected, if the technique offers more than one result.

Estimates. This section presents the selected estimates and a possible range of variation of the selected levels. It could include a statement as to whether the estimates should be considered as maximum or minimum levels.
Component projection of the population

This technical section is different from the previous ones because of the assumptions needed for projecting the individual components of population change.

Projection of the components. This section should start with a brief presentation of the information accepted for the base year of the population projection. It discusses the base population, mortality, fertility, and migration. It should include references to previous sections where such variables were fully analyzed. Next, it should explain how each of the components of population growth was projected.

1. If the projected levels were extrapolated, information on both the data and the technique used for the extrapolation should be presented.

2. If the projected levels were based on assumptions, such assumptions should be clearly specified in the report. In addition, if an interpolation was made between the levels for the base and end-year of the projection, the method of interpolation should be explained.

3. This section should describe how the age and sex patterns of the projected levels of mortality, fertility, and migration were derived.

4. The accepted levels and sex-age patterns should be presented in appendix tables.

If alternate assumptions were made for any of the component variables, they also should be presented.

Projection of the population. This section should indicate which population projection program was used. It should include results for selected years of the projection period and indicate where to obtain detailed tables of the population projections (such as single ages, special age groups, or single calendar years). If more than one population projection series is provided, this section should indicate which is the recommended series.