The Springer Series on Demographic Methods
and Population Analysis 38

Gordon A. Carmichael

Fundamentals of Demographic Analysis: Concepts, Measures and Methods

Springer

The Springer Series on Demographic Methods and Population Analysis

Volume 38

Series Editor
Kenneth C. Land, Duke University

In recent decades, there has been a rapid development of demographic models and methods and an explosive growth in the range of applications of population analysis. This series seeks to provide a publication outlet both for high-quality textual and expository books on modern techniques of demographic analysis and for works that present exemplary applications of such techniques to various aspects of population analysis.

Topics appropriate for the series include:

- General demographic methods
- Techniques of standardization
- Life table models and methods
- Multistate and multiregional life tables, analyses and projections
- Demographic aspects of biostatistics and epidemiology
- Stable population theory and its extensions
- Methods of indirect estimation
- Stochastic population models
- Event history analysis, duration analysis, and hazard regression models
- Demographic projection methods and population forecasts
- Techniques of applied demographic analysis, regional and local population estimates and projections
- Methods of estimation and projection for business and health care applications
- Methods and estimates for unique populations such as schools and students

Volumes in the series are of interest to researchers, professionals, and students in demography, sociology, economics, statistics, geography and regional science, public health and health care management, epidemiology, biostatistics, actuarial science, business, and related fields.

More information about this series at http://www.springer.com/series/6449

Gordon A. Carmichael

Fundamentals of Demographic Analysis: Concepts, Measures and Methods

Springer

Gordon A. Carmichael
School of Demography
Australian National University
Canberra, ACT
Australia

ISSN 1389-6784 ISSN 2215-1990 (electronic)
The Springer Series on Demographic Methods and Population Analysis
ISBN 978-3-319-23254-6 ISBN 978-3-319-23255-3 (eBook)
DOI 10.1007/978-3-319-23255-3

Library of Congress Control Number: 2015949814

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

This book has its origins in lectures prepared while teaching the course Principles of Population Analysis (PPA) in the Demography Program of the Australian National University (ANU) between 1989 and 1998. After a 1-year break I also taught the course for a further 2 years in 2000 and 2001 after moving to the ANU's National Centre for Epidemiology and Population Health (NCEPH) in early 1999. There it was taught in conjunction with a heavily overlapping NCEPH course titled Population Analysis for Health Research. It seemed to me absurd that two courses with such substantial common content had been taught side by side in the same university, each to limited numbers of students, virtually throughout the 1990s, and my argument that common sense recommended in the future teaching them together was accepted by both my previous (Program) and my new (Centre) directors.

By the middle stages of my time employed by the Demography Program, what began as lecture notes delivered to students orally and via a whiteboard, overhead projector and occasional handouts had morphed into draft chapters of a book that were distributed one by one as the course progressed. August 2001, however, saw the arrival of a new director at NCEPH, and armed with an external review recommending such action, he soon decreed that teaching a coursework master's degree that was regularly attracting low single-figure numbers of students to most of its courses was an inefficient use of resources that could more profitably be redeployed boosting the Centre's research productivity.

My classroom teaching consequently ended, and PPA reverted to being the responsibility of the Demography Program, subsequently to become first the Demography and Sociology Program then the Australian Demographic and Social Research Institute (ADSRI). My successor teaching PPA was a former student, Dr Rebecca Kippen, who sought my permission to continue using my draft chapters as core teaching material (she actually had them spiral bound for distribution as an in-house 'textbook'), and would thereafter encourage me towards formal publication whenever she saw me – with the addition of a chapter on population projections (excluded because projections were taught in a separate dedicated course). Rebecca continued using my 'book' until she left the ANU in early 2010 to take up a future fellowship at the University of Melbourne, but I was unable to justify spending the time needed to bring it to publication while in the employ of NCEPH – a demographic methods text was not core business for a Centre with a public health focus.

Retirement in late 2011 put the project back on my agenda as I moved back to ADSRI as an adjunct (i.e., honorary) associate professor. It was not my first priority given the freedom I now enjoyed to do my own thing – I had a series of papers that had also laid dormant because they were not public health-focused that I wanted to write first. But in late 2013 ‘the book’ rose to the top of the ‘to do’ list, and during 2014 I have extensively reviewed and revised what I initially wrote many years ago and have added the recommended chapter on projections.

The book chapters that developed out of my lecture notes were written in the context of teaching groups of generally around 20 students, most of whom were international students from a variety of developing countries. The master’s programme in demography towards which my teaching was directed was at that time domiciled in the ANU’s Development Studies Centre, which relied heavily on international student fee income to fund the salaries of academics it directly employed. An overriding characteristic of those classes was an enormous range of quantitative aptitudes, with students ranging from a few with first degrees in statistics to larger numbers with only the most rudimentary high school maths training. The task was to pitch my teaching at a level that would cater to this enormous range and in particular not lose those at its lower end. I also needed to contend with the reality that most of my students had only just arrived in Australia, and many had come from cultures where the teacher-student relationship was distinctly hierarchical and were not attuned to questioning and interacting as equals with a lecturer. I wanted them to ask questions and let me know if I was not getting through to them.

A number of strategies were pursued to these ends. One was to take my own tutorials, through which I could ensure via worked examples that key issues and principles were reinforced and could personally field student questions. Another was to offer optional extra tutorials for students who were having difficulty, at which they could identify where their problems lay and set the tutorial agenda accordingly and we could work through issues at their pace. But the major strategies in these pre-‘PowerPoint’ days were to (i) develop teaching material that would obviate the need for students to spend lectures furiously scribbling notes and enable them to instead focus on understanding lecture content and (ii) present that teaching material in a way that was as comprehensible as I could make it for those less at ease with quantitative processes and concepts.

This book is the ultimate product of my efforts to make demographic analysis more accessible to students with limited quantitative backgrounds, without being too pedestrian for those with greater mathematical aptitude. Others will be the judges of how successful I have been. It is intended to be an introductory text, leaving ample scope for resort to additional texts and other sources pitched at higher levels with students for whom such material is appropriate. The book has benefitted from the time I spent teaching two courses as one, in particular through the addition of much of the material in the section ‘Further Issues in the Analysis of Mortality’ at the end of Chap. 4. This was introduced in more embryonic form to my teaching in 2000–2001 because of its relevance to students of public health, but has been substantially expanded in the process of revision and updating engaged in this year. Some of the examples presented in the book may seem a little dated. This is a consequence of their often having been compiled during the 1990s when I originally wrote much of this material. Where it has seemed updating was essential (e.g., where trends were being traced), I have taken the trouble to make the necessary revisions, but if an example was a good illustration of a particular technique or principle, I have not bothered replacing it with a more recent one merely to be more ‘up to date’. That examples also frequently draw on Australian data reflects my having taught in an Australian university; I no more apologize for this than previous authors of similar texts I’ve read have for drawing on British or American examples with which they were familiar.

Some of the material in this book has origins in teaching material prepared some years ago by one-time colleagues Professor Peter McDonald (who taught me as a PhD student in 1978) and the late Dr Alan Gray. I gratefully acknowledge their inspiration.

Canberra, ACT, Australia Gordon A. Carmichael

Contents

Chapter 1
Basic Sources, Concepts, Definitions and Types of Measures

Demography, Demographic Analysis and Population Studies

What is 'demography'? A short answer, sufficient perhaps to quickly satisfy the lay inquirer, is 'the study of human populations'. More elaborate definitions are, however, to be found in the literature. Several are reproduced in Table 1.1, where page references make it clear that it is common for books like this one to commence by addressing the question just posed. Some definitions are more restrictive than others. The Belgian Achille Guillard is credited with having first coined and defined the term 'demography'. Of the more recent definitions cited, that attributed to the IUSSP advances us a little beyond 'the study of human populations', but remains general and non-specific. Hauser and Duncan provide more detail, introducing the notion of 'territorial distribution' to provide a touchpoint with geography and itemizing four processes through which populations change over time. Bogue asserts the 'statistical and mathematical' nature of demography, adds 'marriage' as a fifth process generating change and claims theory-building as a long-run disciplinary goal. Shryock, Siegel and Associates in their encyclopaedic two-volume survey of demographic techniques offer that demography may be defined either 'narrowly or broadly', whilst Wunsch and Termote explicitly acknowledge only three processes of change, although their second sentence broadens the agenda *á la* Shryock et al. Weeks's definition is a real catch-all, that of Hinde highlights future prediction as an important element of demography, Preston, Heuveline and Guillot are succinct (perhaps too succinct to be very helpful), and Weinstein and Pillai emphasize that demographers deal with *aggregates* of *living* individuals. Siegel and Swanson's definition revises that of Shryock, Siegel and Associates in the earlier edition of *The Methods and Materials of Demography*.

Clearly the boundaries of demography are fuzzy. There is, however, general agreement that the discipline has a core, variously referred to as ***demographic analysis*** or ***formal demography***, which is concerned with the measurement of popu-

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_1

Table 1.1 Some definitions of demography

Guillard (1855) cited in Shryock, Siegel and Associates (1973: 2)
[Demography is] the natural and social history of the human species or the mathematical knowledge of populations, of their general changes, and of their physical, civil, intellectual, and moral condition.
International Union for the Scientific Study of Population (1958: 3)
Demography is the scientific study of human populations, primarily with respect to their size, their structure and their development.
Hauser and Duncan (1959: 31)
Demography is the study of the size, territorial distribution, and the composition of population, changes therein, and the components of such changes, which may be identified as natality, mortality, territorial movement (migration), and social mobility (change of status).
Bogue (1969: 1–2)
Demography is the statistical and mathematical study of the size, composition, and spatial distribution of human populations, and of changes in these aspects through the operation of the five processes of fertility, mortality, marriage, migration and social mobility. … [Its] long-run goal is to develop a body of theory to explain the events that it charts and compares.
Shryock, Siegel and Associates (1973: 2)
Demography is the science of population [and] may be defined narrowly or broadly. The narrowest sense … [deals] with the size, distribution, structure, and change of populations. … A broader sense includes … ethnic, social, and economic characteristics. … The widest sense … extends to … problems … related to demographic processes: the pressure of populations upon resources, depopulation, family limitation, eugenics, the assimilation of immigrants, urban problems, legislative apportionment, manpower, and the maldistribution of income.
Wunsch and Termote (1978: 1)
Demography is the study of population, its increase through births and immigration, and its decrease through deaths and emigration. … demography is also the study of the various determinants of population change and of the impact of population on the world around us.
Weeks (1994: 4)
Demography is concerned with virtually everything that influences, or can be influenced by population size, distribution, processes, structure, or characteristics.
Hinde (1998: 1)
Demography is the study of population structure and change. With the increasing complexity of modern society, it is becoming ever more important … to make estimates of … the future size and composition of the population. Demographers … carry out this task.
Preston et al. (2001: Preface)
[D]emography [is] the study of population processes. [It studies] the behavior of human populations.
Weinstein and Pillai (2001: 5)
Demography [is] the science of population … its chief concern is with the characteristics of *aggregates* of living individuals. … They are born, they age, they move from place to place, they reproduce, they become ill, and they die.
Siegel and Swanson (2004: 1)
Demography is the scientific study of human population, including its size, distribution, composition, and the factors that determine changes [therein]. [It] focuses on five aspects of human population: (1) size, (2) distribution, (3) composition, (4) population dynamics, and (5) socioeconomic determinants and consequences of population change.

lation phenomena, and a broader interpretative component, often termed ***population studies***, which seeks to understand and explain the patterns, differentials and trends revealed by demographic analysis and to assess their implications and linkages to other phenomena. Population studies tends to focus on relationships between demographic and non-demographic variables, and to the extent that demographers' views as to the scope of their discipline vary, they vary largely in their perceptions of the breadth of the field of population studies. You may also encounter the terms ***mathematical demography***, ***social demography*** and ***anthropological demography***. The former refers to the application of formal mathematics to the analysis of demographic change and relationships; the latter two terms to study of the social and cultural underpinnings of demographic levels, trends and differentials.

Population studies is unambiguously interdisciplinary, overlapping with fields like sociology, anthropology, psychology, economics, geography, history, epidemiology and medicine. Indeed many specialists in population studies develop their interests following initial trainings in one of these other fields, and claim allegiance to that field in addition to their allegiance to population studies. Many academic demographers are in fact employed by university Schools or Departments of Sociology or Geography. But demographic analysis is also interdisciplinary, having obvious overlap with mathematics and statistics and sharing analytic techniques with the likes of biology, epidemiology and medicine. The strongly interdisciplinary nature of both population studies and demographic analysis makes demography, as the discipline that embraces both of them, a particularly interesting, fertile and useful (i.e., policy-relevant) area of academic and intellectual endeavour.

This is a book about demographic analysis; about the technical core of demography upon which population studies depends. A prerequisite to demographic analysis is demographic data: that is, counts of the populations of geographic areas; data on people's demographic characteristics (age, sex, marital status, labour force status, etc.); and counts of demographic events (births, deaths, marriages, migratory moves, etc.). We will now examine the major sources of such data, and the strengths and drawbacks of each.

Sources of Demographic Data

There are four major sources of demographic data: population censuses, sample surveys, vital (sometimes termed 'civil') registration systems which in a few countries are embedded in more extensive population registers, and a variety of administrative systems which yield demographic data as by-products of the administrative functions they serve.

Population Censuses

Population censuses, usually although not always also incorporating housing censuses, are conducted with varying degrees of regularity in most countries. According to the United Nations Statistics Division (2013), countries and territories that failed to conduct a population census during the decade 1995–2004 were Angola, Burundi, Cameroon, Chad, DR Congo, Djibouti, Eritrea, Ethiopia, Guinea-Bissau, Liberia, Madagascar, Mayotte, Nigeria, Somalia, Sudan, Togo, Western Sahara, El Salvador, Colombia, Peru, Afghanistan, Bhutan, DPR Korea, Lebanon, Myanmar, Uzbekistan and Bosnia-Herzegovina. As of mid-2013 seven countries had yet to set a census date for the decade 2005–2014 – Eritrea, Somalia, Western Sahara, Iraq, Lebanon, Pakistan and Uzbekistan. Another 17 had set only a vague date (e.g., 2013) that suggested they were in grave danger of not holding a census, or had passed a nominated date without a census having occurred – Benin, Central African Republic, Comoros Islands, DR Congo, Equatorial Guinea, Gabon, Gambia, Guinea, Madagascar, Sierra Leone, Guatemala, Haiti, Honduras, Afghanistan, Syria, FYR Macedonia and Ukraine. From these lists it is easy to appreciate that among the forces that discourage census taking are smallness, poverty and internal political strife. Small countries sometimes have alternative sources of demographic data, or are so sparsely populated that there is limited demand for such data. Governments of poor countries may simply have other priorities for expenditure of limited economic resources and may lack the administrative infrastructure to support a census. Strife-torn areas tend to present insurmountable practical obstacles to census taking, in addition to which the population may be abnormally mobile, substantially geographically displaced and have sections intent on not being located by government representatives. The government in turn may have other priorities in terms of expenditure and occupying the time of civil servants. There is, in addition to countries like those listed above, a range of other, mostly European, countries that also do not conduct full censuses because they maintain population registers that either on their own, or in conjunction with other sources such as sample surveys, yield equivalent data. These types of sources are dealt with later in this chapter.

A census aims at ***complete coverage*** of a population, although sometimes an element of sampling intrudes, with some respondents being asked a broader range of questions than others. Information is sought about each individual present within a defined geographic area at a defined point in time, with some attempt perhaps also being made to obtain data on persons ‘usually resident’ in the area but temporarily absent from it. Population censuses have several significant disadvantages. They are ***expensive***, which along with the long lead-time needed to plan them and the time required to process and disseminate data after collection limits their frequency. Australia is one of a minority of countries (Canada, New Zealand, Japan, South Korea, Hong Kong SAR, and several small Pacific nations are others) to conduct censuses as often as every 5 years, although even in these countries the unexpected can intervene. The scheduled 2011 New Zealand Census, for example, was deferred for 2 years because the Christchurch earthquake occurred a fortnight before census night, devastating the offices from which it was being administered. The more usual intercensal interval is of the order of a decade, which means that the ***timeliness*** of census data is frequently less than ideal. Some countries attempt to address this problem by conducting intercensal surveys, but while these can pinpoint major changes at national and regional levels they tend not to meet the needs of local area planners and those interested in small minority groups within the population.

The expense of population censuses contributes as well to their characteristically ***limited content***. The huge number of respondents means that the number of questions must be kept in check so as to contain costs of stationery, printing, coding, data capture and, if the census is not self-enumerated, interviewing. Modern computer technologies, such as the Intelligent Forms Processing (IFP) software used for the 2001, 2006 and 2011 Australian Censuses, can facilitate major savings on data processing in countries equipped to use them (Australia also made savings in 2011 through 28 % online submission of census forms, up from 9 % in 2006, the first year with an 'e-census' option), but the case for restricting census content relies only partly, and secondarily, on cost. Censuses are often ***self-enumerated***; the questionnaire is dropped off, the respondent or a household representative on behalf of all household members answers the questions, and a fieldworker collects the completed form. Given this system, gathering data of acceptable quality necessitates ***limiting respondent burden***, so as not to give rise to unacceptable levels of non-response or inaccurate response. An important element of respondent burden is the ***length of the questionnaire***. If perceived to be too long or too time-consuming respondents may avoid answering it altogether, stop answering it part way through, or become careless or even frivolous with their answers. A second important element is the ***complexity of questions asked***. Especially when designing a self-completion census it is essential to appreciate the diverse intellectual capabilities to be catered to, and to keep questions simple so that they can be coped with at the lower end of this intellectual range. Inability to understand a question obviously adversely affects response to that question, but may also contribute to a rapid drop in a person's enthusiasm for cooperating with the census in a more general sense. Thus, demographers and other users have to accept that a self-enumerated census is not a vehicle through which to ask questions of the complexity and depth one might ask in an interviewer-administered sample survey. ***Nor is it a vehicle through which to ask sensitive questions.*** Again, data quality at the levels of both individual topic and the census as a whole can be undermined by questions which are perceived to be offensive, intrusive, or to rekindle unpleasant memories (e.g., a question on the number of children ever born to women aged 15 or older often asked in Australian censuses has always explicitly targeted live births only, in recognition that to ask respondents to include stillbirths, or imply that they should be included, could be distressing).

As a result of the considerations just discussed, population censuses ***typically ask a limited range of core questions*** that enable the basic structure and composition of national, regional and other geographically defined populations to be established. Attributes covered in most censuses include age, sex, marital status, birthplace and/or ethnic group, level of education and occupation. Other questions frequently aim at capturing household composition and/or family type, childbearing experience, religion, income, and any recent change in place of usual residence (typically within the last year and/or the last 5 years). In countries without reliable vital registration systems, considerable emphasis may be placed on censuses gaining insight into fertility and mortality levels and trends. In general, they have been less effective in providing data on mortality than on fertility. Death is a sensitive topic, and for that reason alone difficult to collect direct information on. It also not infrequently leads to the breakup of a household, survivors dispersing to other locations, including other countries. This may lead to a death being attributed to the wrong (regional) population, to its not being reported at all, or to multiple reporting where the responsibility for reporting is unclear. In addition there may be no survivor to report a death; or defective recall may create uncertainty over whether such an event occurred within a nominated reference period (e.g., the preceding 12 months). There are, however, less direct census-based approaches to estimating mortality levels available. One involves comparing the reported age-sex structure of the population with those of what are known as model stable populations, while estimates of child and adult mortality can respectively be made by questioning mothers about the survival of their children, and people about the survival of their parents or first spouses.

The tenor of this discussion has been rather negative, stressing the limitations of population censuses. Obviously they also have positive attributes, otherwise they would never take place. Above all, ***population censuses are head counts***, and as such have administrative utilities that in politicians' eyes normally outweigh their importance as data sources for demographers. In many democracies the census is the basis upon which electoral boundaries are regularly revised, and simple population counts may play a major role in the distribution of resources among competing geographic areas. In Australia, for instance, constitutional and legislative requirements concerning electoral redistributions and the allocation of Federal tax revenues (in particular GST consumption tax revenues) to State and Territory Governments are major reasons for conducting population censuses, especially as regularly as 5 years apart.

The complete coverage aimed at by population censuses means that, provided they come close to achieving that objective, ***reliable data can be obtained for small geographical areas and small subgroups of the population***. Small-area data are indispensable to effective local area planning (determining where to locate or close down schools, where the demand for various community services is concentrated, etc.), and provide crucial input to the location decisions and marketing strategies of businesses. Indeed, data at small-area and higher subnational levels of geographic aggregation are the 'bread and butter' of ***applied demography***, a phrase coined to encompass analyses of demographic data that are intended first and foremost to inform decisionmaking. The ability to focus on small subgroups (e.g., small immigrant and ethnic groups) is a feature of population censuses that is very attractive to academic demographers as well as to planners. These are groups that in any general sample survey of modest size would not be sufficiently well represented for reliable conclusions about them to be drawn, and whose whereabouts often cannot otherwise be adequately pinpointed for more specialized survey work. Potentially, an attractive feature of censuses is their capacity to provide sampling frames for such survey work, but in practice the need to strictly enforce confidentiality assurances in order not to jeopardize the overall integrity of the census tends to make authorities extremely reluctant, if not totally unwilling, to permit this sort of application of census data, especially by non-government agencies.

Complete coverage also means that ***detailed cross-tabulations can be prepared which are not subject to sampling error***. Moreover, in the event that some questions in a census are not asked universally, the size of the sample typically is so large that relevant data items are subject to very low sampling errors.

Population censuses are ***the primary source of the denominators required by a host of demographic rates and ratios***. Countries with reliable vital registration systems could not optimally exploit them without regular censuses, many demographic indices combining a ‘number of events’ numerator from a vital register with a ‘population at risk’ denominator from, or based upon, a census. Often these indices can be estimated using survey data as well, the survey yielding both numerator and denominator. But the ideal is reliable registration data combined with appropriate census data. For census years the census may directly yield denominators, perhaps with a small adjustment depending on the timing of the census during the year, since denominators usually need to be ***mid-year*** populations at risk. For other years denominators normally have a prior or subsequent census base, adjusted for vital and other (especially migration) events that modified that base in the interim.

Finally, successive censuses are invaluable for ***the historical record of a population they build up, the capacity they provide for focusing on demographic change, and the insights into the nature of demographic processes which studies of intercensal change can provide***. Censuses as a historical record obviously have broad academic interest, as Fig. 1.1, showing census-based age-sex pyramids for Australia between 1861 and 2011, attests (those before 1911 amalgamate colonial census data for New South Wales, Victoria, Queensland, South Australia, Western Australia and Tasmania). ***Age-sex pyramids*** are a graphical device in which males are conventionally represented on the left and females on the right, with age groups (typically, but not necessarily, 5-year age groups) ranging from the youngest at the pyramid base to the oldest at its apex. Pyramids shown in Fig. 1.1 have male and female surpluses in age groups distinctively shaded for emphasis. This assemblage of age-sex pyramids captures the development of a population that initially bears the imprint of male-dominated gold rush immigration during the 1850s through a structure still displaying remnants of that era but also, at younger ages, evidence of a more settled, but high fertility, population in 1881 to one that in 1891 had experienced a new influx of immigrants aged in their late twenties and early thirties. Undercutting of the base of the 1901 pyramid highlights depression-induced fertility decline through the 1890s, while by 1911 the base is broadening again, signifying recovery. The 1933 pyramid then reflects the effect on fertility of rampant unemployment during the Great Depression following the Wall Street stockmarket crash of October 1929. By this time the male excesses of population at most ages

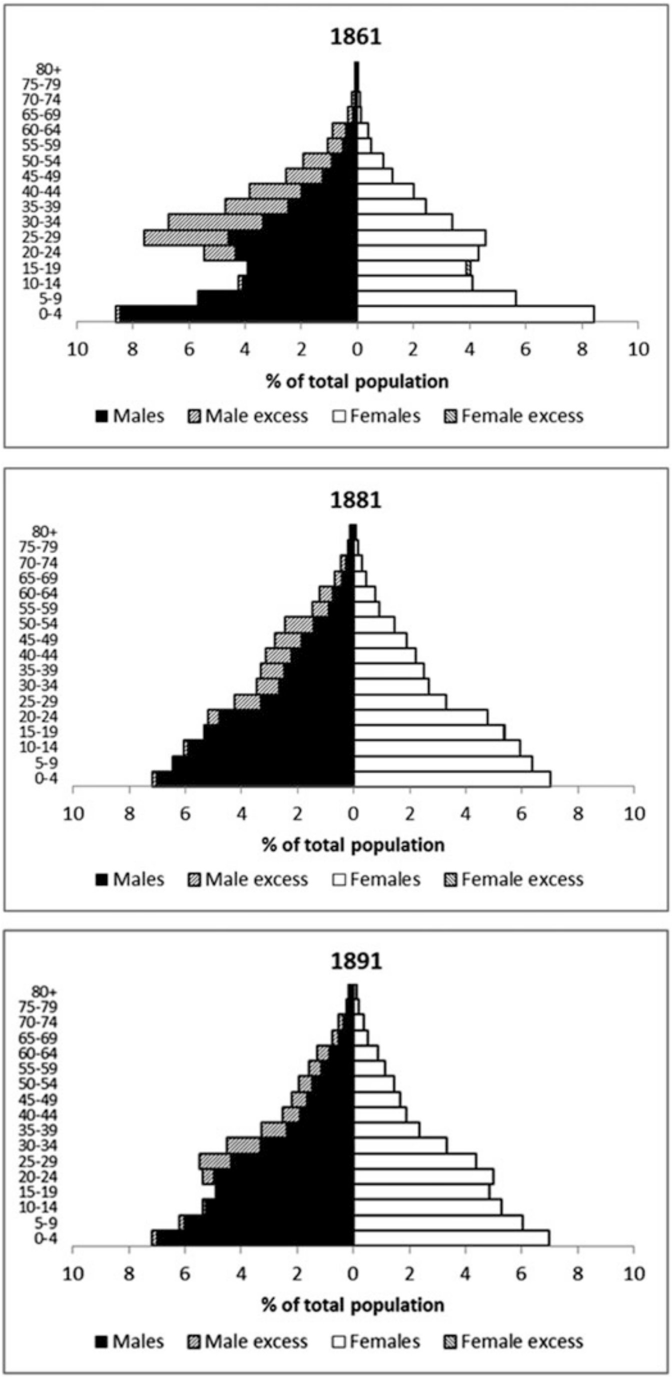


Fig. 1.1 Age-sex pyramids for Australia: selected censuses 1861–2011

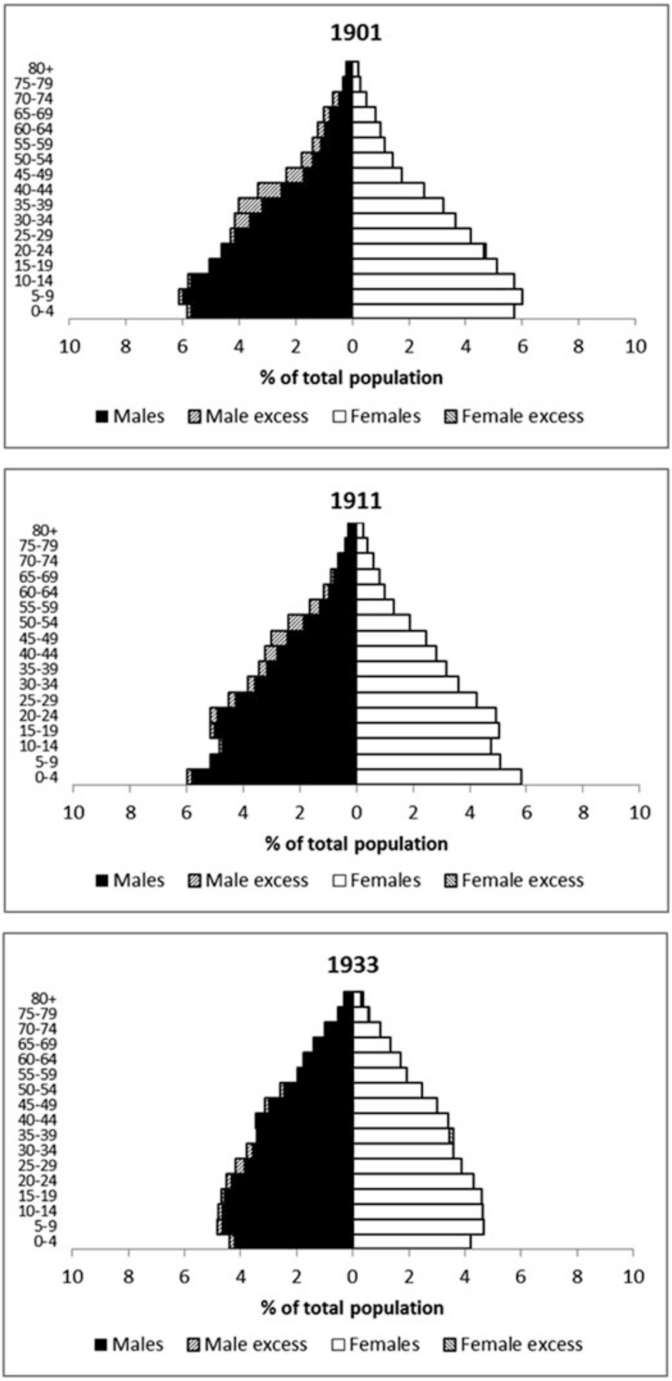


Fig. 1.1 (continued)

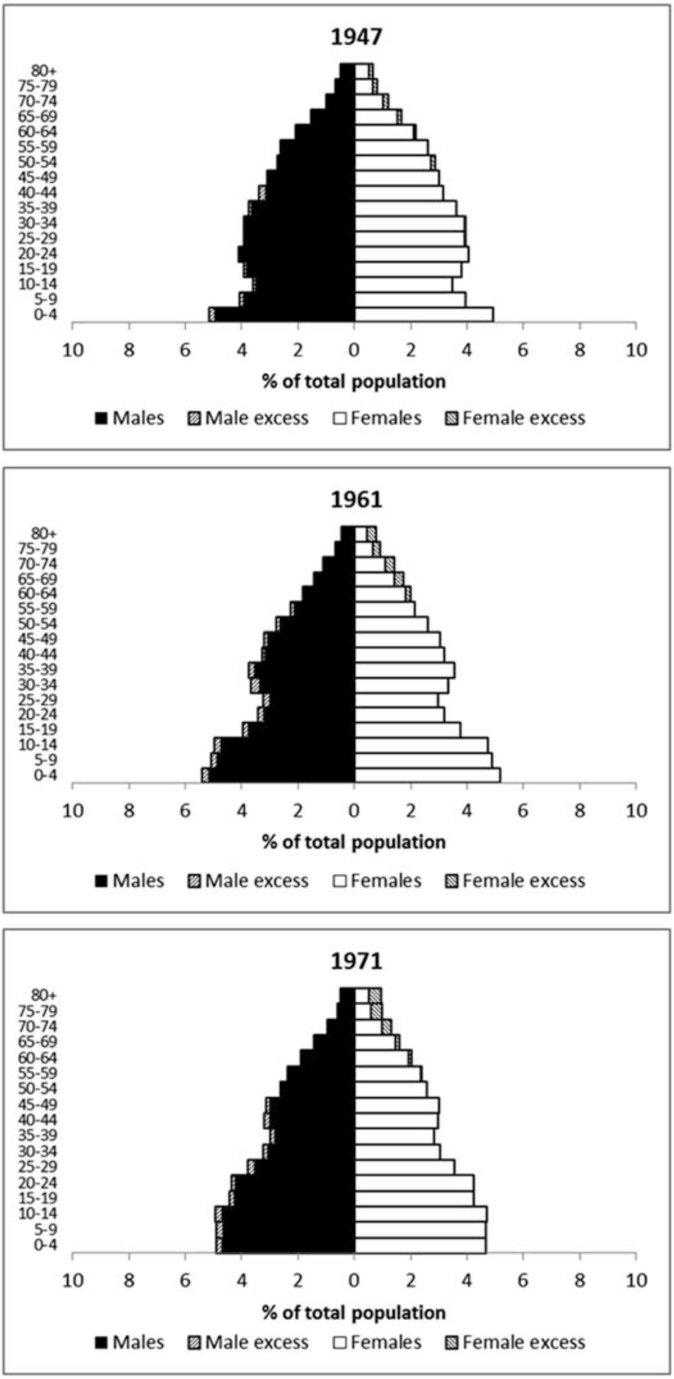


Fig. 1.1 (continued)

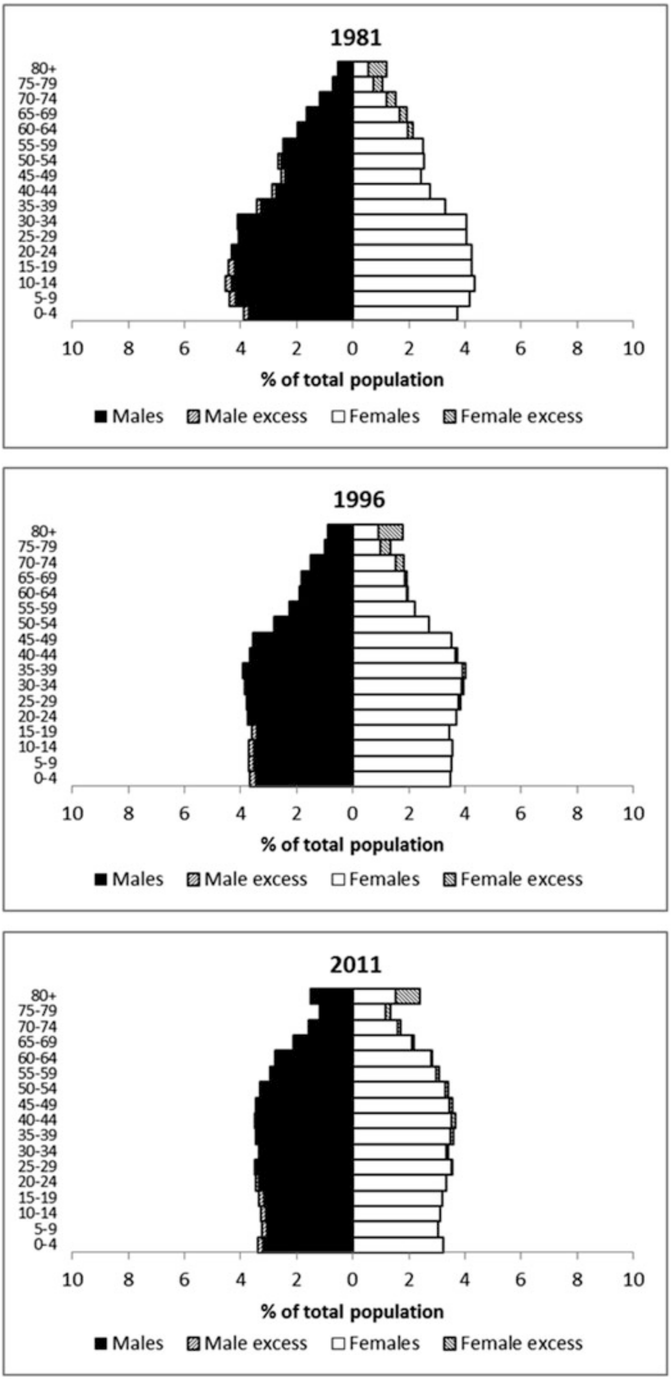


Fig. 1.1 (continued)

have also become much more muted, the heavily male-dominated immigrant cohorts of the nineteenth century having largely died out. Age-sex pyramids for 1947 and 1961 are then notable for three features: development of the post-war Baby Boom bulge; the re-emergence due to immigration of a male surplus at young adult ages; and the appearance of a marked excess of females at older ages as female longevity increased more rapidly than did that of males. By 1971 the first hint of renewed fertility decline appears at the pyramid base, oral contraception having by then been available for a decade, and this has become more pronounced by 1981. Note also that male excesses, reflecting the biological reality of a male-dominated sex ratio at birth, continue to prevail at childhood ages. By 1996 the Baby Boomers are in their thirties and forties, the pyramid has a narrower base, and female population excess at the oldest ages has become more pronounced. Ageing of the population is even more marked by 2011, by which time female population excesses are evident at all ages 30–34 years and above. There is also the hint of a new fertility resurgence, possibly temporary and reflecting the short-term impact of pronatalist government policies dating from 2004.

The historical record censuses provide may also have very practical significance, for example in allowing the demographic evolution of small geographic areas to be traced with a view to predicting future change and planning accordingly. Some suburbs within cities, for example, are initially rapidly settled by young families and age through subsequent censuses with those families. Others maintain more stable demographic profiles through constant turnover of population in well defined categories (single young adults; young families occupying rental accommodation prior to acquiring their own homes; etc.). Still others suddenly change character through gentrification, invasion by minority groups, etc. Each of these scenarios has distinctive planning implications.

As data from a new census become available a priority of many users is to establish change since the previous census. Often there is strong suspicion of change in a particular direction, the census being relied upon to confirm that suspicion and measure the extent of the change. The use of census data to study change ***ideally requires that at each date censuses be based on the same or equivalent questions, coding schemes and field procedures.*** For those who design censuses the lesson here is ***to carefully evaluate the impact on data comparability*** of making alterations in these areas. If an adverse impact is envisaged, advantages expected in terms of better quality data on a topic, cost savings, etc. should be weighed against the loss of temporal comparability. For ***users*** of census data the lesson is to always check for intercensal changes to question format, coding conventions and field procedures. Otherwise it is easy to find oneself inventing fantastic explanations for spurious trends.

Sample Surveys

A sample survey involves questioning a small, ideally ***randomly selected***, sample of a population, although samples may also be ***stratified*** to overrepresent groups (strata) that otherwise would not have sufficient numbers for separate analysis. Biases introduced to the sample by stratification can be offset by ***weighting*** survey data. Surveys may be ***cross-sectional*** (i.e., relate to a set point in time), ***longitudinal*** (reinterview the same respondents at several points in time), ***retrospective*** (collect data about past events – life histories), or ***prospective*** (record information about events as they happen over time). They often focus on only parts of a population (e.g., persons aged 65 and over; women of reproductive age (15–49)), and this ***capacity to focus on population subgroups of special interest*** is a major attraction. It allows questionnaire design to concentrate on probing a subgroup's character, behaviour and attitudes in-depth. Another attraction of surveys relative to censuses is their ***comparative cheapness***, making them affordable for agencies besides those of national governments.

Unless mounted by government agencies empowered to compel response, sample surveys ***can be biased by non-response***. Among those who do respond, however, ***data quality is easier to control*** in surveys than in censuses. Higher data quality is not, though, automatic; indeed survey data quality is extremely variable, and it is vital to always check and evaluate quality control procedures followed before analysing a survey dataset. But whereas censuses frequently rely on self-enumeration, surveys, while they can also use self-enumeration, including these days online enumeration, more often than not use face-to-face or telephone interviews. If interviewers are well trained (and assessing this is part of an evaluation of quality control procedures), they should be able to establish a rapport which maximises the chance of a respondent treating the questionnaire seriously, guide respondents to correct interpretations of questions, probe for accurate answers to questions respondents are unsure of the answers to, recognize and seek resolution of inconsistencies in answers to related questions, and so on. Because of this quality control it is common for surveys to in fact be used to check the accuracy of census and even vital registration data. Thus census agencies often run ***'post-enumeration' surveys*** with this objective in mind, and indeed may use surveys to test proposed new questions before deciding whether or not to include them in a census.

As a corollary largely of the quality control trained interviewers bring to a sample survey, but also of the capacity to define one's target group and topic more narrowly, surveys ***allow much more detailed data to be gathered*** than do censuses. Two of the better known demographic surveys (or, more strictly, groups of surveys) have been the World Fertility Surveys (WFS) and the Demographic and Health Surveys (DHS). The former (Cleland and Scott 1987) saw comparable sample surveys of women aged 15–49 carried out in 66 countries between 1973 and 1984. Most (44) were developing countries, but several large countries (India, China, Brazil and the Soviet Union) declined to participate. Very detailed information was collected on pregnancy histories, the survival of children, knowledge and practice of contraception, and potentially related background social and economic variables. The DHS program, initiated in 1984 and continuing to this day, has sponsored around 300 surveys in over 90 developing countries to collect data on fertility, reproductive and maternal health and mortality, child health, survival and immunization, HIV/AIDS, malaria and nutrition. Most have targeted women or ever married women aged 15–49, although a few, chiefly in Africa, have also targeted husbands. With both the WFS and the DHS considerable resources have been devoted to developing, evaluating and improving survey methodologies and instruments, with a view to producing high quality data. Another survey program focused on providing comparative data on demographic change, this time in developed countries since the mid-1960s (the 'second demographic transition' as it is often referred to), was the series of Fertility and Family Surveys (FFS) sponsored by the United Nations Economic Commission for Europe. Designed to facilitate analysis of recently emergent demographic trends using techniques known as 'event history analysis', these surveys were carried out in 24 chiefly European countries during 1988–1999 (Canada, New Zealand and the USA also participated) (Festy and Prioux 2002). Individual countries (especially more developed countries) also have their own regular surveys that gather demographic data (often in conjunction with labour force data – e.g., the monthly Current Population Survey run since the early 1940s by the US Bureau of the Census for the Bureau of Labor Statistics), and over time there have been myriad one-off or longitudinal demographic surveys, or surveys with content of interest to demographers, mounted by university researchers, specialized research agencies, government departments, etc.

An increasingly important variant of the sample survey in demography over recent decades has been ***the 'micro' approach***. It is generally applied in small geographic areas (often a village) over an extended period of time, combining the formalities of localized census-taking and demographic surveying with detailed semi-structured or unstructured anthropological or sociological investigations (so-called ***'qualitative' research***). The 'micro' approach is more a tool of population studies than of formal demography; a means of shedding light on the complex social and cultural determinants of demographic variables and change. In one sense it can be a troublesome approach. Whereas with conventional survey research there is typically a questionnaire and a description of sampling and field procedures that can be scrutinized and critically evaluated, 'micro' studies do not inherently offer opportunity for such checks. Considerable faith must be placed in the integrity of the fieldworker, who while perhaps wishing to test hypotheses should not have embarked on the project intent on verifying preconceived ideas. With a semi-structured or unstructured survey instrument and the capacity to quote informants selectively, it is relatively easy to tell, and difficult to refute, any story one cares to tell. Nevertheless, in the hands of an honest, skilled practitioner the 'micro' approach is a very valuable tool of demographic explanation. It is, incidentally, useful to distinguish between ***'respondents'*** (those who answer structured survey and census questionnaires where question formats remain unchanged from person to person) and ***'informants'*** (those who respond to more conversational semi-structured or unstructured questioning in in-depth interviews, which tend to be topic-driven and sufficiently flexible to allow questioning to digress to follow up leads as they present during the interview).

Vital Registration Systems

Vital registration systems, sometimes termed ***'civil'*** registration systems, exist more often and tend to be better developed in more developed countries. They derive, as their name implies, from requirements that designated members of the population register ***vital events*** (births, deaths and marriages), usually within some prescribed period after their occurrence. In Australia, responsibility for registration rests with a parent in the case of births and the marriage celebrant in the case of marriages. Registration of deaths is more complex, being 'based on information supplied by a relative or another person acquainted with the deceased, or an official of the institution where the death occurred and on information supplied by a medical practitioner or a coroner as to the cause of death' (Australian Bureau of Statistics 2007: 82).

While demographers might like to imagine otherwise, vital registers, especially in developed countries, ***generally exist primarily for legal and administrative reasons*** rather than to facilitate demographic research. They establish certain facts about individuals, and are used for purposes such as providing documentary proof of age, settling deceased estates and inheritance entitlements, and establishing obligations to support children and entitlement to marry. Their utility as data sources is considerable, since they normally record not merely the occurrence of a demographic event but certain attributes of the event and of the person(s) experiencing it. However, that utility is apt to be viewed by those who maintain vital registers as of secondary importance to their legal and administrative utility. In Australia the situation is complicated by registration being the responsibility of six State and two Territory governments, each with its own registration forms, and there has, for example, been less than universal enthusiasm for amending birth registration forms to eliminate a serious defect in data on mothers' parities (i.e., the numbers of live births they have had) (Carmichael 1986; Corr and Kippen 2006). This defect arises from forms historically requesting details only of previous children *'of the current marriage',* and more recently only of previous children *'of the current relationship'*, whence women with children who repartner and have further children can in doing so revert to parities previously attained but since passed beyond (e.g., a woman with two children who divorces, remarries, then has a third child statistically has a second 'first' birth, not a third birth). From 2007 the Australian Bureau of Statistics began publishing parity (or 'previous issue') data based on children of *all* relationships for those States and Territories now collecting such data, but Victoria and Queensland, between them accounting for around 45 % of national population, still refuse to do so. The central issue is one of sensitivity – whether a woman should be required to disclose the existence of a previous child of which her current husband or partner may be unaware. In the 1960s when non-marital births often led to placement for adoption this was a real issue, but with such births these days typically occurring to cohabiting couples and placements for adoption by strangers uncommon, the sensitivity argument has worn thin.

Like censuses, ***vital registration systems are costly***, and this expense is a major deterrent to their establishment in low-income countries. In this case the cost has two dimensions: the need to maintain a permanent infrastructure for an ongoing data collection operation; and the need for a widely dispersed network of registration points readily accessible to all members of a population. Cost is not, however, the only impediment to establishing viable vital registration systems in developing countries. Even where these do exist, ***coverage is often incomplete***. Several factors may be responsible for this: lack of capacity to enforce a legal obligation to register events (it is difficult to expect compliance if failure to comply is rarely detected and/or attracts no sanction); lack of a popular tradition of registration (in many developed countries registration at the behest of the state followed lengthy periods when the Church performed a similar function); and the absence of a comprehensive system of state-funded entitlements, establishing eligibility for which acts as an incentive to register vital events in many developed countries.

While in many countries the problems of cost and achieving reasonable coverage have meant that no attempt has been made to establish vital registration systems, others have chosen to tackle the task gradually by setting up systems province by province or district by district. India, on the other hand, has set up systems in a series of sample areas, and uses the data generated as the basis for national fertility and mortality estimates. It is not necessary for registration of vital events to be complete before it provides usable fertility and mortality estimates; techniques have been developed for adjusting incomplete data (provided they are not *too* incomplete) to, for example, yield estimates of adult mortality. It is, however, important to always assess the completeness of data from a vital registration system before proceeding with analysis. Blindly assuming complete registration when it is seriously *in*complete obviously can lead to totally erroneous results and conclusions.

Vital registration data ***typically are used in conjunction with census data***. They provide the numerator for a demographic rate or ratio while a census provides the denominator. As already noted, though, the latter may have had to be updated for the period between the census and the date to which the calculation pertains, and this updating of population estimates following a census is in itself another common application of vital registration data. Other data may also feature in such exercises, most commonly migration data. These may derive from border control documentation in the case of national population estimates, with data on internal migration being needed in the case of subnational population estimates. For further detail on vital registration systems see United Nations (2001).

Administrative Systems

Data sources falling under this heading ***exist primarily for administrative purposes but have demographic utility as a by-product of those purposes***. It could be argued that this description applies to vital registers, but they are such a fundamental source of demographic data as to be regarded as a separate category. The administrative priority which attaches to what might be termed 'miscellaneous' administrative sources can sometimes mean that access to them is hard to obtain or carries restrictions, and hence their demographic potential is difficult to realize. It can also mean that items of considerable demographic interest either are not collected or are vulnerable to being discarded when data collection procedures and instruments are reviewed.

Data on divorce are in some jurisdictions another type of vital registration data. Divorce is certainly a vital event, but it is not one that normally is registered in the sense that births, marriages and deaths are. Rather, divorce data tend to emanate from the legal process through which those seeking to formally end marriages pass. They are a by-product of the legal system, courts rather than registrars of vital events furnishing statistical agencies with the returns from which statistics are compiled, and so are considered here to be one of the more important miscellaneous administrative sources. In Australia they provide a good example of the vulnerability of such sources, from demographers' perspective, to administrative whim. In 1995 a revision of the form on which Australian divorce data are transferred to the Australian Bureau of Statistics deleted, as a cost-saving measure, the two items giving parties' marital statuses at marriage (whether never married, divorced or widowed). In making this decision the review eliminated researchers' capacity to distinguish dissolutions of first marriages from those of remarriages following divorce and widowhood, a fundamental distinction to be made in any thorough analysis of divorce. Studies invariably show divorce to be more common in marriages where one or both parties has previously been divorced than in those that are first marriages for both parties.

Briefly mentioned in passing earlier as obviating the need for censuses in some countries, ***universal population registers*** are another administrative source of demographic data (Verhoef and van de Kaa 1987). They are essentially extensions of, and incorporate, vital registers, although these exist as separate entities for each type of vital event whereas in a population register they are brought together and amalgamated. 'The population register is a mechanism for the continuous recording of selected information pertaining to each member of the resident population of a country or area, making it possible to determine up-to-date information about the size and characteristics of the population at selected points in time' (United Nations 2001: 75). Like vital registration systems, population registers have legislated underpinnings to create and enshrine the administrative systems that maintain them, require citizens and others to provide information they rely on in a timely way, and perhaps allow certain information to be imported into them from other more specialized databases. They start from essentially a census base – an inventory of inhabitants of an area and their demographic characteristics, including their residential locations – then update individual records as subsequent events and changes of status are experienced, add new ones to cover births and new immigrants, and terminate or suspend records of deceased persons and emigrants. This process includes recording non-temporary changes of residential address, so one important advantage of population registers over vital registers is their capacity to provide data on migration. Changes of occupation may also be recorded, allowing occupational mobility and labour force composition to be monitored.

Because they entail setting up a record with a unique personal identification number for each individual at creation of the register, birth or on joining the population as an immigrant, population registers are ***bothersome at one level in representing a major, ongoing bureaucratic intrusion into the lives of citizens; in having considerable potential for data to be used for purposes not initially advised or intended; and in being potentially tools of repression.*** Indeed, these sorts of issues go a long way to explaining why they are not more widespread. The Australian Government, for example, proposed in the mid-1980s the establishment of a national identity card system that would have had the potential to become a population register, but was ultimately dissuaded from proceeding by strong opposition to its 'big brother' overtones. From the demographer's viewpoint, however, provided the data are accessible, population registers offer attractive potential for both up-to-date cross-sectional and longitudinal analysis. They are well developed in Nordic countries – especially Sweden, Denmark, Norway and Finland – but can raise interesting issues when attempts are made to interrogate them to generate census-type data. In Norway, for example, considerable difficulty was experienced in this context in trying to define 'households', because notified addresses in multi-unit apartment blocks often failed to identify the particular apartment a person lived in. This necessitated creation of a whole new numbering system for apartments in such complexes (Trude Lappegård, personal communication, July 2013).

The precise content of population registers varies from country to country, and these differences dictate the capacity to produce census-type data exclusively from interrogation of the register. Other countries use a mixture of data from a register and other sources, including sample surveys, to produce census-like statistical snapshots of their populations. Indeed, providing sampling frames for more targeted surveys is another useful attribute of population registers. According to the United Nations, countries producing 'census' data exclusively from 'administrative registers' during the decade 2005–2014 were Andorra, Austria, Denmark, Finland, the Holy See, Monaco, Norway, Slovenia, Sweden, Bahrain and Singapore. Those using a combination of such registers and other sources (chiefly sample surveys) were Belgium, the Czech Republic, Estonia, Germany, Iceland, Italy, Latvia, Lichtenstein, Lithuania, the Netherlands, Poland, Spain, Switzerland and Turkey. Population registers are therefore primarily a European data source.

A third commonly used administrative data source is ***records of international movements of population*** derived from border control procedures. Most countries have laws governing who is and isn't eligible to enter the country and under what conditions, and some also restrict departures of their citizens. In implementing these laws information is gathered from persons arriving at, and departing from, recognized points of entry and exit. The coverage of these data varies from country to country as a function of (i) the level of incentive for illegal migration, (ii) governments' capacities to patrol national borders, and (iii) the existence of populations whose traditional lands straddle national boundaries. For countries like Australia and New Zealand, surrounded by water, with international travel occurring through small numbers of airports and seaports, and with coastal surveillance adequate to intercept most attempts at illegal entry elsewhere in the country, coverage is extremely good. But for a country like the USA, with a long common border with a neighbouring country (Mexico) whose residents it mostly seeks to exclude, yet whose much lower standard of living creates a strong incentive to migrate, coverage may be much poorer.

Not that completeness of coverage is the only consideration in assessing the quality of border control data. As a comparison of Australian and New Zealand data on population flows between the two countries has shown (Carmichael 1993), differences in questions asked and in classification concepts and data processing conventions used can have non-trivial effects on estimates of the sizes of migrant, as distinct from visitor, flows between pairs of countries. A phenomenon known as 'category jumping', in which individuals, through deliberate deception or because expectations on arrival or departure change, are allocated to movement categories which ultimately prove to be incorrect, may also have a major impact on data quality. Accurately separating migrants from visitors in border control data is especially important these days, given the hugely increased volumes of international visitor movement since the advent of international air travel. Border control databases have in fact in recent decades often been overwhelmed by volumes of visitor movement, resulting in Australia and New Zealand, for example, deciding to process data for only samples of those deemed to be involved in such movement.

Records associated with the provision of medical and welfare services are another useful category of administrative data. Their more obvious potential is for informing studies of fertility, mortality, morbidity, the aged, etc., but if recording changes of address they may also be a source of data on internal migration. In Australia data on internal migration have been gathered through the quinquennial censuses since 1971, but between censuses, quarterly estimates of interstate migration were until 1986 based on changes of address notified to the Department of Social Security by recipients of the then universal Family Allowance, while since that time they have been based on changes of address advised to the Health Insurance Commission as administrator of the national health insurance system, Medicare. Other administrative systems which record changes of address on a fairly universal basis, such as ***those operated by telephone or electricity companies***, are also, if accessible, potential sources of data on internal migration in the absence of other sources, although the trend from landline to mobile phones with a wider array of providers constantly urging customers to switch allegiance has undermined the use of telephone records for this purpose.

Data Quality

Demographic data are frequently defective. They may be ***incomplete***, as when there is significant non-response to a census or survey, underregistration of vital events, or an administrative source is only partial in its coverage. They may be ***out of date***, whence they should not be discussed in a manner that pretends otherwise and, if the aim is to assess the current situation, explicit attention should be paid to possible changes since the date to which the data pertain. They may also ***contain measurement errors***, which may arise at any stage of data collection and processing; from faulty questionnaire design, defective sampling, inadequate training and/or performance of field staff, lack of knowledge or defective recall on the part of respondents, substandard coding, errors in computer processing, etc.

It is important that a demographer always be alert to the possibility that data may be defective in some way, particularly (though by no means exclusively) when the population concerned is a developing country population. Before undertaking analysis of any dataset you should ***evaluate*** it with a view to identifying any shortcomings or idiosyncracies that might have implications for your intended analysis. Some types of problem have in the past proved so common that checking for their existence should be routine – age misstatement and underregistration of vital events are two examples (see, for example, United Nations (1955)). Age misstatement may in particular be an issue if you are intent on studying age in single-year-of-age detail. You should also, in the course of analysis, ***be awake to unanticipated results***, the first reaction to which should be to ponder whether any problem with your data or their processing could be responsible. Certainly do not embark on dreaming up fantastic explanations for unexpected results until you are absolutely satisfied the results are genuine. And remember, the fact that data exist in a computerized form, or are neatly printed in an official statistical publication, does not render them automatically correct. In evaluating data it is always healthy to be ***slightly suspicious rather than blindly trusting.***

Just because data are discovered to be defective does not make them automatically unusable. In extreme cases a decision to jettison an entire dataset may be justified, but there exist a battery of techniques for handling deficient data. The well known *United Nations Manual X*, for example, is described online as ‘an aid to demographers and population experts to carry out the best possible evaluation and exploitation of data sources, especially those that are incomplete or deficient’ (United Nations 1983). Moreover, defective data problems often can be dealt with adequately by incorporating appropriate caveats into discussions of results of data analyses, or by demonstrating that a realistic tolerance for data error has no substantive effect on your conclusions.

Absolute Measures: The Population Balancing Equation

Demographers deal with two broad types of measure – ***absolute*** measures and ***relative*** measures. Absolute measures in demography are ***counts of people, or of events that occur to people***. The population balancing equation (also known as the 'basic demographic equation'), a simple accounting equation that expresses population change through time, makes use of such measures. It takes one of two forms, depending on whether the population in question is 'closed' or 'open'.

A closed population experiences no migration, either inward or outward. Its size therefore changes as a function of births and deaths alone. With interplanetary travel not yet an option the obvious example is the world's population, although populations defined by attributes that exist at birth and thereafter are unalterable (e.g., the Australia-born population) are also closed (albeit that if geographically scattered they may be difficult to obtain data on), and some national populations (e.g., until the late 1980s, those of many Soviet bloc countries, and these days that of a country like North Korea) can be considered to be virtually closed. ***An open population***, by contrast, is one whose size is affected by migration as well as by births and deaths. Most of the populations demographers deal with are, of course, in this category.

For a closed population the population balancing equation is:

$$\mathrm{P_2 = P_1 + B - D} \tag{1.1}$$

Where $\mathrm{P_2}$ = population at time 2; $\mathrm{P_1}$ = population at time 1; B = births between time 1 and time 2; D = deaths between time 1 and time 2.

The change in population between time 1 and time 2 is then given by:

$$\mathrm{dP = P_2 - P_1 = B - D} \tag{1.2}$$

For an open population the population balancing equation is:

$$\mathrm{P_2 = P_1 + B - D + I - O} \tag{1.3}$$

Where $\mathrm{P_2}$, $\mathrm{P_1}$, B and D have the same meanings as in Eq. 1.1; I = in-migrants arriving between time 1 and time 2; O = out-migrants departing between time 1 and time 2.

The change in population between time 1 and time 2 is then given by:

$$\mathrm{dP = P_2 - P_1 = B - D + I - O} \tag{1.4}$$

The righthand side of Eq. 1.4 adds together two ***components of population change*** which have special names. The component (B – D) is known as ***natural increase***, and the component (I – O) as ***net migration***. The sizes of open populations are subject to change through both natural increase (which may, of course, be negative – i.e., natural ***decrease*** – if deaths over the period in question outnumber births) and net migration; the sizes of closed populations are altered by natural increase alone.

The population balancing equation can in theory be rearranged to make any unknown element its subject, so as to enable that element to be calculated. Its most common application is, however, in the ***estimation of net migration***, especially for subnational geographic units (for which usually there are no migration data deriving from border control procedures) over periods between censuses. It thus often underpins studies of the contribution of migration to intercensal population change at the state/regional/district/suburban level. This type of study produces estimates of net migration which measure the combined impact of ***international*** (across national boundaries) and ***internal*** (across subnational boundaries) migration. If we denote the (I – O) element in Eq. 1.3 by M (standing for 'net migration') we can rewrite the equation in the form:

$$M = (P_2 - P_1) - (B - D) \tag{1.5}$$

In other words, net migration between time 1 and time 2 is total population change, less natural increase. In applying the population balancing equation, and in particular this variant of it, it is important to be aware of whether the population counts P_1 and P_2 are *de facto* or *de jure* counts.

A ***de facto population count*** counts ***all persons*** present in the area of interest (country, region, etc.) at the time the count is made, and counts them ***at the geographic location where they are found*** at that time. It excludes usual residents of the area who are temporarily absent, but includes visitors who are not usual residents. A *de facto* census thus enumerates everybody present at the time of the census according to their geographic location at that time. A ***de jure population count***, on the other hand, counts persons ***usually resident*** in the area of interest, and counts them ***according to their geographic location of usual residence***. It includes usual residents who are temporarily absent (to the extent that they can be traced) but excludes visitors who are not usual residents. A *de jure* census thus enumerates the usually resident population according to individuals' usual places of residence.

It is not always clearcut what constitutes being a 'usual resident' or a 'visitor'. Arbitrary definitions based on duration of residence are sometimes imposed, but it is more common to rely on self-classification by respondents – they are allocated to whichever category, 'usual resident' or 'visitor', they perceive themselves to belong to, and usual residents are also left to indicate whether their address on census night is their 'usual' address. Note, too, that in distinguishing between *de facto* and *de jure* censuses we are ***making a distinction at two levels*** about any individual: first, ***whether*** (s)he should be counted at all; and second, assuming (s)he should be counted, ***where***, geographically, (s)he should be counted.

De facto censuses and surveys can cause problems if one's interest is household composition or family type. Temporary absentees or visitors can distort household and family structures. At the 1981 Australian Census, for example, use of a strictly *de facto* count resulted in families with one parent temporarily absent on business etc. on census night being classified as 'single parent' families. Addition of a question on persons temporarily absent from the household at the 1986 Census introduced a *de jure* element to counter this ludicrous misclassification. This was an example of a data problem being identified because of increased focus on a particular population subgroup. Single parent families became of much greater interest to Australian researchers during the 1970s after (i) the divorce rate rose substantially following mid-decade liberalization of divorce laws and (ii) the introduction of a Supporting Mother's Benefit in 1973 made choosing to retain children born outside marriage rather than place them for adoption more economically feasible.

We have strayed from the population balancing equation, but the point is this. When applying Eq. 1.3 in respect of P_1 and P_2 values which are *de facto* population counts, I and O refer to ***all*** arrivals and departures respectively. When applying it in respect of *de jure* population counts, I and O refer to ***permanent*** arrivals and departures ***only*** (i.e., to those which add people to, or remove them from, the 'usual resident' population; the word 'permanent' is an overstatement for some individuals, but is intended to encompass all who, in crossing the relevant geographic boundary, change their place of usual residence). Similarly, if using Eq. 1.5 to compute net migration, whether the result gives net ***total*** or net ***permanent*** migration depends on whether P_1 and P_2 are *de facto* or *de jure* population counts respectively. ***Never mix the two!***

Relative Measures: Rates and Probabilities

Absolute measures such as the population balancing equation both produces as output and requires as input are of limited use to demographers, except as input into the calculation of more sophisticated measures. Suppose, for example, we are told that in Great Britain there were 606,000 deaths in 1871 and 560,000 deaths in 1921. If we subtract the later figure from the earlier one we discover that there were 46,000 fewer deaths in 1921 than in 1871, and 46,000 is 7.6 % of 606,000. But can we conclude that mortality in Great Britain declined by 7.6 % during 1871–1921? The answer is an emphatic 'No!' Why? Because ***the populations at risk*** of dying were not the same in 1871 and 1921.

In 1871 the population of Great Britain was 27.4 million; by 1921 it had increased to 44.0 million. Suppose this population had not changed, and had still been 27.4 million in 1921. What effect would we expect this to have had on the number of deaths in 1921? Obviously we would expect considerably fewer than 560,000 deaths. If a population of 44.0 million produced 560,000 deaths, then a much smaller population (27.4 million) experiencing the same mortality conditions would be expected to yield many fewer deaths, widening the gap between the numbers of deaths in 1871 and 1921. Thus we can conclude that the absolute decline in the number of deaths understates the extent of the mortality decline. Had the population of Great Britain remained constant at 27.4 million a much larger decline in the number of deaths would have occurred.

To make comparisons between populations (whether two or more populations at one point in time or, as in the example above, one population at two or more points in time) we need measures of the ***relative*** occurrence of demographic events; measures which allow (or control) for differences in the sizes of the populations producing those events. Always be sceptical of analyses of demographic data that rely exclusively on absolute numbers. Journalists seeking to quickly write up the latest press release from the national statistical agency or to localize the focus of demographic data to the area their publication serves are serial offenders in this matter.

Demographic Rates

Two major categories of measures of relative occurrence used by demographers are ***rates*** and ***probabilities***. The same two categories are sometimes referred to as ***m-type rates*** and ***q-type rates*** (Hinde 1998), the ‘m’ and ‘q’ being standard life table notations (see Chap. 4). Common use of the word ‘rates’ in this dichotomy is an unhelpful source of confusion, but in this book the concept ‘rate’ will be rather strictly defined, and the distinction between it and the concept ‘probability’ will be observed.

Both rates and probabilities are examples of ***ratios***, the name given to any expression of the form ‘a divided by b’, in which the ‘a’ quantity is known as the ***numerator*** and the ‘b’ quantity as the ***denominator***. Even aside from the confusion just discussed, the term ‘rate’ is used very loosely in the demographic literature, and you should be aware that some measures that conventionally are called ‘rates’ do not conform with the strict definition to be presented here. Two prominent examples are the infant mortality rate (IMR), which is not a ‘rate’, but an estimate of a probability, and the total fertility rate (TFR), which is not a single rate but the sum of a series of age-specific rates. In some quarters the practice has developed of calling the TFR the total fertility ***ratio***, but even that is not an accurate description, leading the prominent British demographic journal *Population Studies* to insist that authors refer simply to ‘total fertility’. Some indices that are no more than simple percentages are also conventionally called ‘rates’; for example, the ‘literacy rate’ – the percentage of a population which is literate.

A demographic rate can be defined as ***the number of occurrences of a given type of demographic event*** (birth, first birth, death, first marriage, etc.) ***in a given period of time*** (usually a calendar year, or a 1-year period of cohort experience) ***divided by the average size of the population at risk of experiencing the event during the given time period***. Strictly speaking, the denominators of rates should be ***the number of person-years of exposure to risk***. This concept is discussed later in this chapter, where it is shown that the average, or mean population at risk is ordinarily a satisfactory approximation of the number of person-years of exposure to risk. As the former is invariably far more easily calculated it tends to be the denominator used for practical purposes. There are various ways of calculating the mean population at risk; for example, averaging the populations at risk at the beginning and end of ***the reference year*** (i.e., the year for which the rate is to be calculated), or averaging the populations at risk at the beginning of that year and at quarterly intervals thereafter. It is also common to use ***the mid-period (mid-year) population at risk*** as an estimate of the mean population at risk, mid-year population estimates often being routinely produced by national statistical agencies. This can lead to slight variations in results if events modifying the population at risk over time are abnormally clustered in one part of the reference year. Having said that, however, use of a mid-year, compared to a mean, population at risk normally has very little impact on a rate's value, even under conditions of quite marked clustering.

Suppose, for example, we have the following data for a closed population in which a disease epidemic occurred during the final months of 2001. We want to calculate a crude death rate, being the ratio of deaths in 2001 to the mean population at risk of dying during that year.

Population at 31 December 2000	1,050,850
Deaths in first quarter 2001	5,400
Deaths in second quarter 2001	6,200
Deaths in third quarter 2001	5,800
Deaths in fourth quarter 2001	15,200
Live births in first quarter 2001	10,200
Live births in second quarter 2001	10,400
Live births in third quarter 2001	10,600
Live births in fourth quarter 2001	10,300

We have a marked clustering of deaths in the final quarter of 2001, which might lead us to suspect that it would be inadvisable to use the mid-year population at risk as an estimate of the mean population at risk. Since we are dealing with a closed population we can obtain populations at the end of each quarter by taking the 31 December 2000 population and successively adding live births and subtracting deaths for each quarter. This process yields the following populations:

Population at 31 December 2000	1,050,850
Population at 31 March, 2001	1,055,650
Population at 30 June 2001 (mid-year)	1,059,850
Population at 30 September 2001	1,064,650
Population at 31 December 2001	1,059,750

The mean population at risk can be estimated by averaging these figures, since they are equally spaced through the reference year. The result (1,058,150) indeed differs from the mid-year population at risk (which is 1,059,850). However, if we divide total deaths during 2001 (32,600) by each of these figures and multiply by 1,000 (to obtain a rate per 1,000 at risk) we get answers of 30.8 per 1,000 and 30.8 per 1,000. In other words, even with a marked clustering of events modifying the population at risk, use of the mid-year population at risk to approximate the mean population at risk had no impact on the value of a crude death rate calculated to one decimal place. Thus, we can almost always use a mid-year population at risk as the denominator of a demographic rate. ***As a general rule, demographic rates are far more sensitive to changes in their numerators than to changes in their denominators***. The latter can tolerate some approximation, hence inaccuracy, without results being significantly affected. On the other hand, changes in rates over short time periods are normally attributable mainly to changes in their numerators.

We will return now to our example of mortality in Great Britain in 1871 and 1921. In place of the absolute numbers of deaths with which we began we will compute relative measures by calculating crude death rates. The crude death rate (CDR) is given by:

$$\mathrm{CDR} = (\mathrm{D}/\mathrm{P})\,.1{,}000 \tag{1.6}$$

Where D = deaths during year y; P = mean (or mid-year) total population in year y

Substituting the data for Great Britain in this equation (assuming the populations given to be mid-year populations) we have:

$$\mathrm{CDR}_{1871} = (606{,}000/27{,}400{,}000)\ 1{,}000 = 22.1 \text{ deaths per 1,000 total population}$$
$$\mathrm{CDR}_{1921} = (560{,}000/44{,}000{,}000)\ 1{,}000 = 12.7 \text{ deaths per 1,000 total population}$$

This establishes a decline of $22.1 - 12.7 = 9.4$ deaths per 1,000 total population, or 42.5 %, between 1871 and 1921, a far more spectacular decline than the 7.6 % decline in total deaths established earlier.

In this example, measures of absolute and relative occurrence do at least both show declines. It is, however, possible for them to change in opposite directions. Suppose, for example, there had been 660,000 deaths in Great Britain in 1921 instead of 560,000. This would have been 54,000 ***more*** deaths than in 1871, but would have yielded a CDR of 15.0 per 1,000 total population, still appreciably ***below*** the 1871 figure. Thus, despite the absolute incidence of mortality having increased, the relative incidence would have declined, the latter being the more meaningful finding because of its controlling for change in the population at risk.

The CDR is an example of a ***true rate***. Its denominator accurately captures the population at risk of experiencing the event (death) that is the focus of its numerator. There are other so-called ‘crude’ rates that represent looser uses of the term ‘rate’. While they also employ the mean (or mid-year) total population as denominator, not all persons in this purported population at risk are *actually* at risk of experiencing the event focused on in the numerator. Examples are the crude birth rate (CBR) and crude marriage rate (CMR), which are calculated as follows:

$$\mathrm{CBR} = (\mathrm{B}/\mathrm{P})\,.1{,}000 \tag{1.7}$$

$$\mathrm{CMR} = (\mathrm{M}/\mathrm{P})\,.1{,}000 \tag{1.8}$$

Where B = live births during year y; M = marriages during year y; P = mean (or mid-year) total population in year y.

While all persons in a population are at risk of dying over any given period, not all are at risk of giving birth or getting married. Males, women yet to reach menarche (the onset of first menstruation, and hence the biological ability to bear children) and women past menopause (the cessation of menstruation, and the biological ability to bear children) are not at risk of giving birth. Likewise, persons who are legally too young to marry or are currently married are not (in a monogamous society) at risk of getting married. Obviously neither the CBR nor the CMR is a true rate, because in neither case does the denominator conform with the definition of a rate given above. It is not hard to see, either, why these 'rates' are called 'crude'.

But what about the CDR? As a 'true' rate, why is it also called a 'crude' rate? It is crude for another reason; it does not allow for extreme variability in the risk of dying at different ages or for the concentration of high risk at one end (the older end) of the age distribution. This means that CDRs can differ, or change, simply because the age structures of the relevant populations differ or change, without there necessarily being any difference/change in the ***underlying*** risks of dying at different ages. In practice, of course, differences/changes in CDRs typically reflect differences/changes in both age structure ***and*** underlying risks of dying; but the CDRs themselves tell us nothing of the relative strengths of these forces or, indeed, whether both are operating in the same direction. As an example to be presented shortly will show, it is not uncommon for one CDR to be higher than another, but for the differential in underlying ***age-specific*** risks of dying to run in the opposite direction. In such a situation, differences in age structure have both compensated for, and outweighed, those in underlying risks of dying, creating the false impression that mortality is higher in the population in which, age for age, it is actually lower.

Those other crude rates that are not 'true' rates are also vulnerable to differences/changes in age structure (and other dimensions of ***population composition***), and are therefore in a sense doubly crude. But so extreme are age differentials in the risk of dying, and so concentrated is high risk at the oldest ages, that the vulnerability of the CDR is acute. Whereas a ranking of national populations by crude ***birth*** rate usually conforms quite well with one based on a more refined measure of fertility, a ranking by crude ***death*** rate, despite the CDR being a 'true' rate, may deviate appreciably from one based on a mortality index which controls for differences in age structure.

Demographic Probabilities

Differentiating probabilities from rates gets confusing when common usage can see one type of measure called the other (the example of the infant mortality 'rate' was mentioned above). Basically, a demographic probability measures the risk of experiencing a specified type of demographic event ***during a given stage of a person's life, or a given life cycle phase***. While the use of estimation procedures may mean that in practice there is not always strict conformity with the statement, a probability measures risk in the context of one's being a member of a particular ***cohort***; a group of individuals ***who experienced some previous demographic event*** (most commonly birth – whence we speak of a 'birth cohort') ***during a specified period of time*** (most commonly a particular calendar year). Rates may also be computed for cohorts, but a cohort setting is not conceptually as integral.

Usually 'stage of life' or 'life cycle phase' is defined in terms of time elapsed since the event conferring membership of the cohort was experienced. If that event is birth, time elapsed since birth is a person's age, and we calculate probabilities of, for example, dying between specified birthdays (or ***exact ages***). But the previous event could also be something like marriage or divorce, whence the life cycle focused on is one spent in the married or divorced state (as compared to the 'alive' state); time elapsed since marriage or divorce is ***duration*** of marriage or divorce; and we calculate, for example, probabilities of divorcing between specified wedding anniversaries (exact durations of marriage), or of remarrying between specified anniversaries of divorce (exact durations of divorce). It is also possible for 'stage of life' or 'life cycle phase' to be defined in terms of calendar dates (e.g., the probability of a person born in year x dying during – i.e., between the beginning and the end of – year y) or even a combination of a calendar date and an exact age or duration. However, most of the probabilities that interest demographers pertain to life cycle phases that begin and end at nominated ***exact*** ages or durations.

Demographic probabilities are distinguished from demographic rates primarily by the nature of their denominators. Both types of measure have numerators that are counts of demographic events during a reference period, or a reference phase of the life cycle, but whereas a rate relates those events to a measure of the average (or mid-period) size of the population at risk, ***a probability can be defined as the number of occurrences of a given type of demographic event during a specified life cycle phase divided by the population at risk at the BEGINNING of that life cycle phase***. Strictly speaking, this definition conforms to a common approach to ***estimating*** probabilities. For refined work adjustments to allow for extraneous demographic processes (migration is a common one) removing people from risk or adding people 'at risk' during the reference phase may be called for. But the definition does capture common practice, just as the earlier definition of a rate did in referring to 'the average size of the population at risk' rather than to the more technically correct 'number of person-years of exposure to risk'.

Consider as an example of a demographic probability the infamous infant mortality 'rate'. The IMR is given by:

$$\mathrm{IMR} = {}_1q_0 = D_0/B \tag{1.9}$$

Where D_0 = deaths at age 0 during year y; B = live births during year y; ${}_1q_0$ is life table notation for the probability of dying at age 0 (this notation is explained in Chap. 3, but is introduced here to reinforce the point that ${}_1q_0$ and the IMR are to all intents and purposes one and the same).

Infant mortality is by definition mortality during the first year of life; that is, mortality between birth (exact age 0) and a child's first birthday (exact age 1). In Eq. 1.9, B is a measure of those entering, or reaching ***the beginning of***, the period of their lives when they would be at risk of infant death during year y. D_0 is an estimate of the deaths during infancy experienced by this group of newborn babies. Thus we have an ***estimate*** of the ***probability*** that a child born in year y dies before reaching its first birthday.

Why is D_0, and hence the IMR, only an ***estimate***? The answer is that, since it takes a year for any child to move from birth to its first birthday, ***the period of risk of infant death for any child invariably straddles two calendar years*** (unless the child happens to be born at midnight on 31st December). Hence, (i) not all of the infant deaths making up our numerator D_0 involved children from our denominator B; some were deaths of children born not in year y, but in year $y - 1$. Similarly, (ii) some of the infant deaths of children in our denominator B will have occurred not during year y, but during year $y + 1$, in which case they are not included in our numerator D_0. Our estimation procedure ***assumes that these two elements of imprecision cancel out***; that infant deaths during year y of children who were born during year $y - 1$ are a good approximation of infant deaths of children born during year y that in fact occur during year $y + 1$.

Assumptions of this type are often made when calculating demographic probabilities. It is common practice to estimate probabilities by dividing the number of people experiencing the ***reference event*** (in the above example, death) at a specified life cycle stage (infancy) ***during a calendar year*** by the population who entered that stage ***in the same year***. The major reason for this practice, which produces only estimates because numerator and denominator do not relate to exactly the same group of individuals, is the reality that statistical agencies tabulate event counts by age/duration and calendar year rather than by age/duration and cohort, the form more suited to calculating probabilities. Using the principles of a graphical device known as the ***Lexis diagram*** (see Chap. 3), demographers frequently convert data from the former to the latter type. But as this, too, is a process of estimation involving assumptions, it is sometimes barely worth the effort. And in the particular case of the IMR, as we shall see, the assumptions underlying Lexis diagram principles are simply untenable.

A straightforward example will illustrate the difference between a rate and a probability in demography. Consider the data over page:

Using Eq. 1.9 we have $\mathrm{IMR}_{1990} = (D_0/B)\ 1{,}000 = (9{,}760/97{,}600)\ 1{,}000 = 100.0$ deaths per 1,000 live births (we have multiplied by 1,000 to obtain an answer in a form comparable to an age-specific death rate). The death rate at age 0 is given

Births in 2005	97,600
Infant deaths in 2005	9,760
Population aged 0 at 30 June 2005	91,800

by (D_0/P_0) 1,000 (where P_0 is the mid-year population aged 0) = (9,760/91,800) 1,000 = 106.3 deaths per 1,000 aged 0. The two measures have identical numerators (albeit that for the IMR this numerator is an estimate of the true numerator), but both conceptually and numerically quite distinct denominators.

A Digression: Concepts of Age, Incidence and Prevalence

You may have noted, in the foregoing few paragraphs, reference to two different concepts of age. Demographers distinguish between ***exact ages*** and ***conventional ages***, otherwise termed ***ages in completed years***, or ***ages last birthday***. For the individual, exact age is ***an instantaneous experience***; no sooner has one attained a nominated exact age than one is older, and no longer that exact age. One is exact age 0 at birth, exact age 1 precisely a year later, and so on. While in countries like Australia birthdays conventionally are celebrated over a full day, in reality one is only ever ***exactly*** x years old for an instant. Moments beforehand one is still younger than x, and moments after one is older (x plus a fraction of a second). By contrast a conventional age, or age in completed years, or age last birthday is an attribute that ***remains with a person for a full year***. One is aged 0 (completed years) for the entire year between birth and one's first birthday; 10 (completed years) for the entire year between one's tenth and eleventh birthdays; and so on.

Demographers often are interested in (i) populations which attained a nominated exact age (reached a particular birthday) during a specified ***period*** of time (often a particular calendar year), and (ii) populations which at a nominated ***point*** in time (say the date of a census, or the mid-point of a calendar year) shared in common a specified age, or series of consecutive ages, in completed years (i.e., were members of a particular ***age group***). Similarly, and by direct analogy, they may be interested in (i) populations which attained a nominated exact ***duration*** (of marriage, divorce, etc.) during a specified period of time, and (ii) populations which at a nominated point in time shared in common a specified duration, or series of consecutive durations (of marriage, divorce, etc.), in completed years. The distinction between these two types of populations should become clearer once Lexis diagrams have been covered in Chap. 3, but it can be used to provide another perspective on the difference between demographic rates and probabilities. ***A demographic rate*** has as its denominator a population quantity based on ages (or durations) ***in completed years. A demographic probability usually*** has as its denominator a population quantity based on ***exact ages (or durations)***.

While strictly an epidemiological rather than a demographic distinction, it is worth also noting at this juncture the difference between measures of ***incidence*** and ***prevalence***. This distinction is made in the context of producing relative measures of the pervasiveness of particular diseases or medical conditions in a population. Its relevance to demographers lies in their interest in understanding the disease patterns and trends that underpin mortality patterns and trends. Incidence measures capture ***the number of new occurrences of a disease or medical condition in a population over a defined reference period*** (typically a calendar year), ***relative to the size of the population at risk*** (typically the mid-year population at risk). Such measures are rates, strictly defined, and thus appropriately referred to as ***incidence rates***. Their numerators are counts of ***events*** (new occurrences, or diagnoses, of diseases or medical conditions).

Prevalence measures, in contrast, focus not on a reference period, but on a point in time (a date), and capture ***the number of cases of a disease or medical condition that exist in a population at that reference date, relative to the size of the population at risk***. Strictly speaking, such a measure is a measure of ***point prevalence***. Particularly when dealing with medical conditions of typically short duration (e.g., the common cold), you may also encounter measures of ***period prevalence***. In this case the numerator is a count not of people with a disease or condition ***at a point in time***, but of those who had that condition ***at any time over a defined period*** (perhaps a month). It is, of course, important not to compare measures of point and period prevalence, since the latter have an inherent upward bias, the degree of which depends on the length of the reference period. In more general terms, though, measures of prevalence are neither rates nor probabilities. They are, in fact, simple ***proportions***. Their numerators are not counts of events, but of persons ***in a particular category*** at a point in time (or at some stage during a defined period). ***When*** the category was entered is irrelevant (it could have been years previously). On the other hand, persons who would feature in the numerator of an incidence rate for the year in which the reference date/period lies, but who at that date/period had already left the category (through death, out-migration or recovery), or were yet to enter it (i.e., would enter it after the reference date/ period), are excluded. Whether incidence, or prevalence measures for a disease/condition tend to be higher depends on its average duration from onset to death or recovery. If disease/condition episodes on average last longer than a year, prevalence will be higher; if on average they are shorter, incidence will be higher. Fairly obviously, ***the two types of measures should not be mixed in comparative analysis***.

Refinement of Demographic Rates and the Effect of Age Structure

We encountered above several so-called ‘crude’ rates: the crude death rate, the crude birth rate and the crude marriage rate. We noted that the latter two measures are not really ‘rates’ at all, because some persons in their denominators are not at risk of giving birth and marrying respectively. As a rule, crude rates do not provide much information about the levels of the demographic phenomena to which they relate; indeed they can be positively misleading. They involve a trade-off between precision on the one hand, and their reliance on relatively unsophisticated data on the other. Their main merit is that they can often be computed in circumstances where more refined measures cannot be; they are better than nothing, although as we shall see they can be so misleading that even that is sometimes debatable.

'Crude' rates deserve their name. Demographers tackle the problem their crudeness poses by seeking to calculate more refined rates. In doing this they aim to ***close in on the population at risk*** by doing one or both of two things: (i) ***eliminating from the denominator persons who could not possibly experience the event in the numerator***; and (ii) ***recognizing that certain personal characteristics make someone more, or less, likely to experience that event.***

In the case of the CBR, a refinement of the former type yields the general fertility rate (GFR). This measure retains live births during year y as its numerator, but its denominator excludes all men, and women who are outside the conventionally defined reproductive ages. Hence we have a 'true' rate:

$$\mathrm{GFR} = \left(\mathrm{B}/\mathrm{P}_{(\mathrm{f},15\text{–}49)}\right).1{,}000 \tag{1.10}$$

Where B = live births during year y; $\mathrm{P}_{(\mathrm{f},15\text{–}49)}$ = mean (or mid-year) female population aged 15–49 in year y.

This measure might also be subjected to a refinement of the second type. In most societies marriage is culturally the most approved setting for bearing children, and women are considerably more likely to give birth if they are married than if they are unmarried. Hence there is a clear case for calculating ***specific*** rates; in this case rates specific for marital status. Specific rates are calculated for ***subgroups*** of the total population at risk by dividing the number of relevant events occurring to that subgroup in the reference year by the mean, or mid-year, population in the subgroup. In this case it makes sense to calculate both a marital fertility rate (MFR) and a non-marital fertility rate (NMFR). These are given by:

$$\mathrm{MFR} = \left(\mathrm{B}_{\mathrm{m}}/\mathrm{P}_{(\mathrm{f},\mathrm{m},15\text{–}49)}\right).1{,}000 \tag{1.11}$$

Where B_{m} = marital live births during year y (live births to married women whose husbands acknowledged paternity); $\mathrm{P}_{(\mathrm{f},\mathrm{m},15\text{–}49)}$ = mean (or mid-year) married female population aged 15–49 in year y.

And:

$$\mathrm{NMFR} = \left(\mathrm{B}_{\mathrm{n}}/\mathrm{P}_{(\mathrm{f},\mathrm{u},15\text{–}49)}\right).1{,}000 \tag{1.12}$$

Where B_{n} = non-marital live births during year y (live births to women who were not married to the fathers of their children); $\mathrm{P}_{(\mathrm{f},\mathrm{u},15\text{–}49)}$ = mean (or mid-year) unmarried female population aged 15–49 in year y.

These two specific rates are less straightforward than most on two counts. First, note that the numerators are not defined simply as ‘live births to married women’ and ‘live births to unmarried women’ during year y respectively. This is because it is possible for a married woman to have a non-marital birth (if the father is someone other than her husband), and for an unmarried woman to have a marital birth (if an event causing her to become unmarried occurred between conception and confinement). Thus, strictly speaking the numerators and denominators do not pertain to the same groups of individuals, although in practice they are usually very close to doing so. Second, who should be considered ‘unmarried’ is not clearcut. Obviously never married, widowed and divorced women qualify, but separated women could be considered either married (which legally they are) or unmarried. Which option one takes depends on whether one judges that births to separated women are more likely to be fathered by their legal husbands (whence they should be recorded as marital births) or by some other partner (whence they would be non-marital births). In Australia the latter may be more likely.

The personal characteristic most commonly affecting a person’s risk of experiencing a particular type of demographic event is age, and the most commonly encountered specific rates are therefore ***age***-specific rates. As a ‘true’ rate, the crude death rate is not a candidate for the type of refinement undertaken in moving from the crude birth rate to the general fertility rate, but it most certainly is a candidate for the type of refinement that is accomplished through the calculation of specific rates. Death is much more likely to occur at some ages than at others, and mortality risks for men and women can also be quite different, both generally and at certain ages (***sex is a second personal characteristic by which demographic analyses are routinely refined***). So variable is the risk of dying by age, in particular, that crude death rates, which take no account of the age structure of a population, can give totally false impressions of the comparative underlying levels of mortality in different populations.

Consider the populations of Malaysia and Australia. In 1988 Malaysia had a CDR of 4.93 deaths per 1,000 mid-year population, while Australia had a CDR of 7.25 deaths per 1,000 mid-year population. These figures suggest that mortality was higher in Australia by a factor of 47 %. Intuitively, to anyone who knows Malaysia and Australia, this sounds peculiar. While Malaysia is hardly among the world’s least developed nations, of the two countries Australia in 1988 was the more developed, and we would expect health services and survival chances to have been superior.

In Table 1.2 the comparison of mortality levels in Malaysia and Australia has been refined through the calculation of age-sex-specific death rates for the two populations. The relevant equation is:

$$\mathrm{ASSDR}_{(s,x)} = \left(\mathrm{D}_{(s,x)}/\mathrm{P}_{(s,x)}\right).1{,}000 \qquad (1.13)$$

Table 1.2 Age-sex-specific death rates for Malaysia and Australia, 1988

	Males		Females	
Age group	Malaysia	Australia	Malaysia	Australia
0	16.6	9.8	12.8	7.6
1–4	1.7	0.5	1.1	0.4
5–9	0.6	0.2	0.5	0.2
10–14	0.6	0.3	0.4	0.2
15–19	1.0	1.1	0.5	0.4
20–24	1.5	1.6	0.6	0.5
25–29	1.7	1.5	0.9	0.5
30–34	2.0	1.4	1.1	0.6
35–39	2.4	1.5	1.5	0.8
40–44	3.4	2.2	2.3	1.2
45–49	5.5	3.4	3.2	2.1
50–54	9.9	6.0	5.9	3.4
55–59	15.4	10.0	9.9	5.5
60–64	25.5	17.3	17.0	8.7
65–69	38.3	27.2	27.8	13.8
70–74	62.7	45.3	46.6	23.5
75–79	86.2	71.9	69.8	40.7
80–84	144.3	110.7	119.0	71.4
85+	173.4	186.6	142.8	147.7

Source: *United Nations Demographic Yearbook, 1990* and Australian Bureau of Statistics, *Deaths Australia 1988*

Where s denotes sex (male or female); x denotes an age group; $D_{(s,x)}$ = deaths of people of sex s at age x during year y; $P_{(s,x)}$ = mean (or mid-year) population of sex s and age x in year y.

A quick inspection of Table 1.2 reveals that for the majority of age-sex groups Australia's age-sex-specific death rate was ***lower*** than Malaysia's, often by a substantial margin. Only for males aged 15–19, 20–24 and 85+, and for females aged 85+, did Australia record a higher death rate. Marginally higher death rates for Australian males in their late teens and early twenties reflect greater susceptibility to involvement in fatal road accidents, while higher Australian death rates for both sexes at ages 85 and over are due to older age structures ***within*** that age group. Of men and women aged 85 and over, proportionately more in Australia than in Malaysia were aged 90 and over, and 95 and over, and therefore subject to the very highest risks of dying. But despite these exceptions, the conclusion invited by Table 1.2 is the reverse of that invited by our earlier comparison of CDRs. The table suggests very clearly that, age for age, mortality was lower, not higher, in Australia.

What causes the CDR for Malaysia to be lower than that for Australia, when for almost all age-sex groups the death rate for Malaysia is higher? The answer is the two populations' different age structures. Higher fertility in Malaysia had given its population in 1988 a younger age structure, and except in the first year of life, age-sex-specific death rates are much lower at younger than at older ages. Malaysia's population was relatively more concentrated at younger ages where the risk of dying is low; Australia's was relatively more concentrated at older ages where the risk of dying is high. The extent of this difference in population composition was more than enough to offset Australia's generally lower age-sex-specific death rates and produce CDRs which convey a totally misleading picture of the comparative ***underlying*** levels of mortality in the two populations.

In case you are still mystified as to how lower age-sex-specific death rates for Australia could produce sufficient deaths to give rise to a higher CDR, Table 1.3 demonstrates what has happened. The table shows how each 1,000,000 members of the populations of Malaysia and Australia in 1988 were distributed by sex and age (for each population, total males + total females = 1,000,000). It also shows the age-sex distributions of deaths occurring in 1988 to the two groups of 1,000,000 people, deaths in each age-sex group having been obtained by applying the relevant age-sex-specific death rate from Table 1.2 to the corresponding number of persons at risk from Table 1.3. We thus have two populations of identical size, each of which has had the appropriate schedule of age-sex-specific death rates applied to it. What

Table 1.3 Age-sex distributions of 1,000,000 people and deaths occurring to those people in Malaysia and Australia, 1988

	Males				Females			
	Malaysia		Australia		Malaysia		Australia	
Age group	Persons	Deaths	Persons	Deaths	Persons	Deaths	Persons	Deaths
0	14,254	237	7,597	74	13,464	172	7,245	55
1–4	56,773	97	30,324	15	53,804	59	28,954	12
5–9	61,207	37	37,762	8	58,012	29	35,803	7
10–14	56,002	34	38,712	12	53,613	21	36,758	7
15–19	52,776	53	43,626	48	50,772	25	41,743	17
20–24	50,193	75	40,771	65	49,060	29	39,297	20
25–29	42,748	73	42,914	64	44,839	40	41,982	21
30–34	35,704	71	40,234	56	38,326	42	39,933	24
35–39	30,427	73	38,778	58	31,269	47	38,391	31
40–44	23,858	81	36,248	80	23,075	53	34,537	41
45–49	20,473	113	27,923	95	19,767	63	26,357	55
50–54	17,052	169	23,855	143	17,190	101	22,798	78
55–59	12,615	194	22,708	227	13,604	135	21,924	121
60–64	10,043	256	21,764	377	10,513	179	22,336	194
65–69	7,373	282	17,701	481	8,498	236	19,973	276
70–74	4,608	289	12,857	582	5,246	244	16,200	381
75–79	3,430	296	8,712	626	3,976	278	12,555	511
80–84	1,399	202	4,471	495	1,675	199	7,827	559
85+	963	167	2,319	433	1,399	200	6,111	903
Total	501,898	2,799	499,276	3,939	498,102	2,152	500,724	3,313

Source: Prepared from data presented in *United Nations Demographic Yearbook, 1990* and Australian Bureau of Statistics, *Deaths Australia 1988* and *Estimated Resident Population by Sex and Age: States and Territories of Australia June 1988 and Preliminary June 1989* do we find? We find that the population with the generally ***lower*** age-sex-specific death rates (Australia) produces the ***larger*** total number of deaths (7,252 compared to 4,951). How has this come about?

If you compare the two distributions of 'Persons' by age and sex in Table 1.3 you will note that Malaysia has substantially larger numbers at younger ages (up to age group 20–24). After that Australia has the larger numbers, its advantage becoming very pronounced beyond about age 60. There are at least twice as many of each 1,000,000 people in Australia in each of these older age-sex groups as in Malaysia, and by ages 80–84 and 85+ the ratio for females is well over four times as many. The older age groups are also where the great majority of deaths are concentrated, and are clearly the source of Australia's higher total number of deaths. Although generally having lower death rates at these ages, ***those rates are applied to significantly larger numbers of people at risk***. In effect, the degree to which Australia's death rates at these ages lie ***below*** those of Malaysia is, by a considerable margin, more than offset by the degree to which its populations at risk ***exceed*** those for Malaysia. The result is very much larger ***absolute numbers*** of deaths at older ages in Australia, purely because the age distribution of the population is biased towards those 'high risk' ages.

In order to sensibly compare mortality levels in different populations we need to eliminate such bias. The CDR does not do this, and is in consequence not only not a very helpful index, but a positively dangerous one. Death rates refined by age (and sex) do eliminate age-bias, and while the added insight gained is generally not as spectacular or necessarily as clearcut as in the example just presented, there is a case for similarly refining almost any summary rate used in studying a demographic phenomenon.

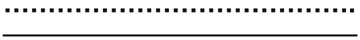
The Concept 'Person-Years Exposed to Risk'

While we have characterized the denominator of a demographic rate as the mean, or mid-period (generally mid-year), population exposed to risk, we also noted that, strictly speaking, this denominator ***should*** be the number of ***person-years*** for which the population at risk was exposed to that risk. Mean and mid-period populations at risk are ***estimates*** of person-years of exposure to risk. But what do we mean by 'person-years exposed to risk'? Essentially we mean ***the total amount of time spent at risk by all the individuals who were at risk of experiencing the event of interest at some stage during the reference period***.

Suppose we wish to compute some rate for calendar year y, and that this rate pertains to a non-renewable event E. A ***non-renewable event*** is one that an individual can experience only once (such as death, first marriage or having a first birth); a ***renewable event*** is one that can recur in an individual's life (such as giving birth, getting married or divorcing). The number of person-years of exposure to the risk of experiencing event E during year y is the sum of two components:

1. Each person who was at risk of experiencing event E at the beginning of year y, but who survived right through that year ***without*** experiencing it and remained at risk at the end of the year, contributes one person-year of exposure to risk.
2. Each person who actually experienced event E during year y, or who only began to be at risk part way through year y, or who ceased to be at risk part way through year y for a reason ***other*** than that (s)he experienced event E (or who fits into more than one of these categories) contributes ***part*** of a person-year of exposure to risk. The fraction of a person-year each contributes equals the fraction of year y spent at risk.

The second of these components clearly has several sub-components. Both they and the first component can be summarized diagrammatically, the solid line indicating the reference period (year y; the period of ***potential*** exposure to risk), the dotted line indicating the period of ***actual*** exposure to risk, and the letters E, J and L standing for 'event', 'join' and 'leave' respectively (i.e., occurrence of the ***event*** of interest; a person ***joining*** the population at risk part way through the reference period; and a person ***leaving*** the population at risk part way through for a reason other than that event E was experienced). ***Component 1***, using these conventions, is represented as follows:



Component 2, subcomponent 1 consists of persons at risk at the beginning of year y, who experienced event E during that year. Each was at risk for the fraction of year y extending from the start of the year to the date when event E occurred. The occurrence of event E terminated their exposure to risk.

Component 2, subcomponent 2 consists of persons who began to be at risk part way through year y, and continued to be at risk for the remainder of the year, not experiencing event E. Each was at risk for the fraction of year y extending from the date of joining the population at risk until the end of the year. Reasons for joining part way through the reference period vary, depending on the rate being calculated. In-migration is almost always a possibility; an in-migrant may have been at risk of event E in some ***other*** population earlier in year y, but only begins to be at risk ***in the population under study*** after joining it. Other reasons might include, in relation respectively to a crude death rate, an age-specific death rate, and a marriage-duration-specific divorce rate, for example, being born, attaining a certain age, or attaining a certain marriage duration during year y. One isn't at risk of dying until born, or of dying at a particular age or divorcing at a particular marriage duration until that age or marriage duration has been reached. In general, then, the occurrence of some event J, causing one to join the population at risk, initiates a person's exposure to risk.

Component 2, subcomponent 3 consists of persons who were at risk at the beginning of year y, but ceased to be at risk during the year for a reason ***other*** than that they experienced event E. Each was at risk for the fraction of a year that had elapsed when the period at risk ended. Out-migration terminates a person's exposure to risk as a member of a population. So does death (even when it is ***not*** event E), and so, too, for example, does attainment of a particular age or duration if it takes a person ***out of*** the age or duration category for which a rate is being calculated. One isn't at risk of anything if dead; neither is one at risk of dying at a given age or divorcing at a given marriage duration once having passed beyond that age or marriage duration. In general, then, the occurrence of some event L, causing one to leave the population at risk, terminates a person's exposure to risk.

Component 2, subcomponent 4 consists of persons who combine elements of the first and second subcomponents. They joined the population at risk part way through year y ***and then*** experienced event E before the end of the year. Each was at risk for the fraction of the year that elapsed between these two occurrences.

Component 2, subcomponents 5 and 6 consist of persons who combine elements of subcomponents 2 and 3. Either they joined the population at risk part way through year y, then left it again before the year ended without experiencing event E, or, having been at risk at the start of year y, they left the population at risk without experiencing event E, but rejoined it later in the year. These two subcomponents can be represented diagrammatically as follows:

It is possible to conceive also of more complex subcomponents entailing multiple entries to and exits from the population at risk, in combination perhaps with ultimate experience of event E, but we will ignore those. All of this probably makes obtaining the denominator for a demographic rate in terms of person-years exposed to risk seem rather complicated. We will try to clarify matters with ***a simple example***.

Suppose we have a population of 100 people at 1 January in year y; that this population is closed (i.e., migration is not a complicating factor); and that 1 person dies and 3 are born in the middle of each month throughout year y. We will also assume that all deaths occurred to people alive at the start of the year (eliminating the complication of some individuals experiencing more than one event). What is the number of person-years of exposure to the risk of dying during year y for this population?

We will note for a start that we began with 100 people at 1 January in year y, but that with 12 deaths and 36 births through the year the population at 31 December in year y was $100 - 12 + 36 = 124$. Having noted this figure, we proceed as follows:

1. Of our original 100 people, 88 survived right through year y (since all 12 deaths occurred to this initial 100). They contributed 88 person-years of exposure to the risk of dying, and this is component 1 of the answer we seek.
2. The 12 who died, one in the middle of each month, survived for 0.5, 1.5, 2.5, ..., 11.5 months. If we add these periods of survival we find that, in total, those who died survived for 72 person-***months***, which divided by 12 comes to 6 person-years. This is component 2, subcomponent 1 of our answer.
3. What about the 36 who were born during year y, three in the middle of each month? They joined the population at risk part way through the year, and as none later died during year y the three born in January each spent 11.5 months at risk, the three born in February 10.5 months, ..., and the three born in December 0.5 months. The total period spent at risk of dying during year y by persons born during the year was therefore $3(11.5 + 10.5 + \dots + 0.5)$ person-months, which is 216 person-months or 18 person-years. This is component 2, subcomponent 2 of our answer.
4. Because of the assumptions we have made (no migration, therefore nobody leaving the population at risk other than by dying, or joining it other than by being born; no individual experiencing more than one event) subcomponents 3–6 of component 2 all contribute zero person-years of exposure to risk.
5. Hence the total number of person-years of exposure to the risk of dying during year y for our population was $88 + 6 + 18 = 112$.

This just happens to be the average of the beginning-of-year and end-of-year populations we noted earlier ($(100 + 124)/2 = 112$). In other words, the number of person-years of exposure to the risk of dying ***is exactly equal to*** the mean population exposed to risk, which is the denominator we have been using for demographic rates until now. ***Why is this so?***

It is so because the events which altered the size of our risk population during year y, the 12 births and 36 deaths, ***were evenly distributed through the year***. This is the situation we ***assume*** prevails when we use the mean, or mid-year, population at risk as the denominator of a rate. If it is a reasonable assumption to make, it is very convenient. It is much easier and requires much less detailed data to calculate the mean, or mid-year, population at risk than to calculate the number of person-years of exposure to risk. It is, of course, possible to conceive of more complicated examples than the one just presented; examples in which the population at risk might be being affected by more than two types of events, including in-migrant arrivals and out-migrant departures. Such complexity is no cause for alarm. The general finding still holds. ***Provided that occurrences of each type of event affecting the size of a population at risk are evenly spread through the reference period for which a rate is to be calculated, the number of person-years of exposure to risk will equal the mean, and the mid-period, population exposed to risk.***

Before proceeding, we will pause to consider the implications of removing one simplifying assumption from the example just discussed. Suppose that instead of all 12 deaths during year y occurring to people alive at the ***beginning*** of the year, one of them involved one of the people born ***during*** the year. Indeed, we will be specific and say that the death in mid-July involved one of the three people born in mid-February.

1. Under the original assumption that nobody born in year y also died in that year, this birth and death involved two people, one of whom was at risk from mid-February to the end of year y (i.e., for 10.5 months) and the other of whom was at risk from the beginning of year y to mid-July (i.e., for 6.5 months). Together these two individuals were at risk for $10.5 + 6.5 = 17$ person-months.
2. Under the new assumption that the birth and death involved the same person, that person was at risk from mid-February to mid-July (i.e., for 5 months). However, because one of the 12 deaths during year y involved someone born during that year, only 11 of the 100 persons alive at the start of year y died during that year. An additional person therefore survived right through year y (i.e., for 12 months). If we add the period at risk for this individual to that for the person who was born in February and died in July we get a total of $12 + 5 = 17$ person-months, ***the same answer as before***.
3. This tells us that while the reality may be that one person can experience more than one event affecting his/her risk status during a year, for the purpose of calculating total person-years of exposure to risk for a population ***we can assume that no person experiences more than one such event and still get the same answer. Errors*** in the periods over which ***individuals*** were exposed to risk flowing from this assumption ***cancel out***.

This finding is very helpful. Recall that the number of person-years exposed to risk for a population has two components, the second having six subcomponents (and then some). Subcomponents 4–6 relate to individuals who experienced ***two*** events affecting their risk status during year y. Rarely are demographic data sufficiently detailed to permit individuals with this sort of complex experience to be identified. The finding we have made tells us, however, that ***for practical purposes we can ignore subcomponents 4–6 and any more complicated subcomponents we***

may care to imagine. We cover them by treating the individuals concerned ***as if*** they belonged to ***two*** (or more) of subcomponents 1–3.

We will also pause here to reflect on the assumption in our discussion that event E was non-renewable. ***What difference does it make if event E is renewable*** (i.e., able to be experienced by an individual more than once)? The point about renewable events is that experiencing them either does not end a person's exposure to risk (as experiencing a non-renewable event does), or alternatively ends it only temporarily. When calculating person-years exposed to risk for a renewable event we need to consider which of these categories the event fits into. The obvious example of a renewable event that does not end exposure to risk is giving birth. While biology dictates that several months must elapse after a woman has given birth before she can do so again, fertility rates which are not parity-specific (parity-specific births – first births, second births, etc. – are ***non***-renewable events) assume that a woman who gives birth is immediately at risk again. In terms of the component model presented above it is therefore not appropriate to terminate the period of exposure to risk at the date event E occurs; component 2, subcomponent 1 of the model becomes irrelevant, and women giving birth are treated as contributing to component 1 (i.e., at risk throughout year y). Examples of renewable events which ***do*** end a person's exposure to risk, but perhaps only temporarily, are out-migration, marriage, divorce, and remarriage following divorce. While a person may out-migrate, marry, divorce, or remarry following divorce more than once, having once done so (s)he must respectively in-migrate, divorce/be widowed, remarry, or be divorced again before being able to experience the event a second time. This inevitably takes time, and the experience of event E is treated as ending exposure to risk. Any renewal of exposure during the reference period is covered under component 2, subcomponent 2 of the model (i.e., the person re***joins*** the population at risk).

Let us now return to the finding from our example that if occurrences of each type of event which can remove someone from, or add someone to, a population at risk are ***evenly spread*** through the reference period, the number of person-years of exposure to risk equals the mean (and mid-period) population exposed to risk. ***In most instances an assumption that events altering risk status are evenly spread through time is sufficiently accurate for a demographic rate to be computed using the mean (or mid-period) population at risk, without it being necessary to compute person-years of exposure to risk.*** This is a very useful principle, because it greatly reduces the time and effort, and the complexity of the data, required to obtain denominators for demographic rates.

To illustrate, let's consider the example of calculating a crude death rate for Australia for 2010. To calculate the number of person-years of exposure to the risk of dying in Australia in 2010 we need to know:

1. The size of the population at risk at the beginning of 2010.
2. The number and timing of deaths during 2010 (since on dying a person ceased to be at risk of death).
3. The number and timing of births during 2010 (since newborn children were at risk of dying from the dates of their births).

4. The number and timing of arrivals and departures by immigrants and emigrants during 2010 (since these groups respectively became at risk on arrival and ceased to be at risk on departure).

These data, with events distributed by quarter (1st quarter = January–March, 2nd quarter = April–June, etc.), are shown below. The timing of events could be given in greater detail, but distributions by quarter are sufficient to indicate quite some deviation from the even patterns through time we assume when using the mean (or mid-period) population exposed to risk as the denominator for our crude death rate.

Population at 1:1:10 = 21,865,623.

Quarter	Deaths	Births	Arrivals	Departures
1	32,689	73,420	119,239	59,380
2	35,694	73,479	88,499	56,491
3	38,439	73,774	109,580	67,804
4	35,808	69,997	107,801	69,406
Total	142,630	290,670	425,119	253,081

While almost certainly some individuals will have experienced more than one of the four types of events affecting the population at risk of dying during 1980, we noted above that we can calculate person-years of exposure to risk assuming that ***nobody*** experiences more than one of these events, confident that errors at the individual level will cancel out. This assumption means we must treat all deaths and departures as having involved persons who were members of the Australian population at the beginning of 1980. Hence we assume that 21,865,623 – 142,630 – 253,081 = 21,469,912 persons survived right through 1980, and component 1 of the required exposure to risk figure is ***21,469,912 person-years***.

Component 2, subcomponent 1 of this figure consists of person-years spent at risk by those who died during 1980 (all of whom, under our assumption, were members of the Australian population at the start of 1980). Making the further assumption that persons who died in any quarter on average died in the ***middle*** of that quarter, we obtain subcomponent 1 as follows:

Quarter	Deaths	Average person-months lived	Total person-months lived
1	32,689	1.5	49,033.5
2	35,694	4.5	160,623.0
3	38,439	7.5	288,292.5
4	35,808	10.5	375,984.0
Total			873,933.0

Dividing 873,933 person-months by 12 gives approximately ***72,828 person-years***.

Component 2, subcomponent 2 consists of person-years spent at risk by those who were born into the Australian population during 1980 or who joined it as immigrants in that year. We assume that births and arrivals of immigrants in any quarter on average occurred in the middle of the quarter again, and obtain subcomponent 2 as follows:

Quarter	Sum of births and arrivals	Average person-months lived	Total person-months lived
1	192,659	10.5	2,022,919.5
2	161,978	7.5	1,214,835.0
3	183,354	4.5	825,093.0
4	177,798	1.5	266,697.0
Total			4,329,544.5

Dividing 4,329,544.5 person-months by 12 gives approximately ***360,795 person-years***.

Component 2, subcomponent 3 consists of person-years spent at risk by those who emigrated during 1980. We assume that departures of emigrants in any quarter on average took place in the middle of the quarter, and obtain subcomponent 3 as follows:

Quarter	Departures	Average person-months lived	Total person-months lived
1	59,380	1.5	89,070.0
2	56,491	4.5	254,209.5
3	67,804	7.5	508,530.0
4	69,406	10.5	728,763.0
Total			1,580,572.5

Dividing 1,580,572.5 person-months by 12 gives approximately ***131,714 person-years***.

Total person-years of exposure to the risk of dying in Australia during 2010 is the sum of the components and subcomponents now obtained; i.e. 21,469,912 + 72,828 + 360,795 + 131,714 = ***22,035,249 person-years***.

What if, instead of basing our calculations on distributions of deaths, births, arrivals and departures by quarter, we had assumed that events of each type were ***evenly distributed*** through 2010? This is, as we have noted, the assumption we would make in using the mean (or mid-year) population at risk as the denominator for our crude death rate. Given our other assumption that no individual experienced more than one event altering his/her risk status, this new assumption amounts to assuming that, ***on average***, everyone who died, was born, arrived as an immigrant or departed as an emigrant during 1980 was at risk of dying in Australia for exactly 6 months. Under this assumption the components and subcomponents of our person-years of exposure to the risk of dying become:

Component 1 = 21,469,912 person-years (unchanged)

Component 2, subcomponent 1 = 142,630/2 = 71,315 person-years (total deaths divided by 2, since every two deaths corresponds to one person-year of exposure to risk).

Component 2, subcomponent 2 = (290,670 + 425,119)/2 = 357,894.5 person-years (total births plus total arrivals divided by 2, since every two births and every two arrivals corresponds to one person-year of exposure to risk).

Component 2, subcomponent 3 = 253,081/2 = 126,540.5 person-years (total departures divided by 2, since every two departures corresponds to one person-year of exposure to risk).

Total person-years of exposure to the risk of dying in Australia in 2010 is the sum of these components and subcomponents, or 21,469,912 + 71,315 + 357,894.5 + 126,540.5 = ***22,025,662 person-years***.

This 'mean population' exposure to risk estimate differs from the 'seasonally adjusted' one calculated using events distributed by quarter, but what happens when we use each estimate as the denominator for our 2010 crude death rate? Our 'seasonally adjusted' CDR is:

$$(142{,}630/22{,}035{,}249)\,1{,}000 = 6.47 \text{ deaths per } 1{,}000$$

Our 'mean population' CDR is:

$$(142{,}630/22{,}025{,}662)\,1{,}000 = 6.48 \text{ deaths per } 1{,}000$$

The increased accuracy we have gained through adjusting seasonally isn't worth the effort. Although we also obtained the denominator for our 'mean population' CDR using the size of the population at the beginning of 2010 and data on deaths, births, arrivals and departures during that year, we needn't have gone to this trouble. We'd have got exactly the same answer by just averaging the populations at risk at the beginning and end of 2010. This would have been a very simple calculation – certainly much simpler than the procedure we went through to get a 'seasonally adjusted' denominator. ***It usually happens that very little error is introduced by using a 'mean population exposed to risk' denominator when calculating a demographic rate.*** Hence demographers normally make do with these (or with mid-year populations exposed to risk as estimates of mean populations exposed to risk), and don't bother computing person-years exposed to risk in intricate detail.

So why bother introducing the concept of person-years of exposure to risk at all? There are at least three reasons why it is an important concept to grasp.

1. ***Theoretically*** the denominator of any demographic rate should be in this form. It just happens that most of the time demographers make do with simple approximations that (i) use less detailed, and importantly more readily available, data, and (ii) save a lot of work for little added precision.
2. The concept of person-years lived is a key one for understanding life tables.

3. There are some circumstances in which the assumption that events occur evenly over time is ***grossly*** inaccurate. In these circumstances, thinking in terms of person-years of exposure to risk assumes more practical significance. The obvious example, which we will dwell on at greater length in Chap. 4, is infant mortality, because deaths during the first year of life are heavily concentrated ***early*** in the year of exposure to risk (i.e., within the first ***month*** of life, and within that month within the first ***week*** of life).

The Principle of Correspondence

When constructing demographic rates and probabilities, a commonsense sort of principle tells us: (i) that if a person is part of the exposed-to-risk population forming the denominator of such a measure, any occurrence of the relevant event to that person during the reference period should be included in the numerator; and (ii) that if an occurrence of the relevant event to a person during the reference period forms part of the numerator, that person should also be included in the exposed-to-risk population forming the denominator. This principle is known as ***the principle of correspondence***. It is so commonsense as to seem obvious, but its application in practice often requires data of a complexity that is seldom available. Hence, it is frequently violated, and approximations are used.

An obvious example arose in our earlier discussion of the infant mortality rate. We noted that the deaths at age 0 during year y forming the numerator did not all occur to the persons born (attaining exact age 0) during year y who formed the denominator. Some occurred to children born the ***previous*** year. Likewise, not all deaths at age 0 of children born during year y (the denominator population) were included in the numerator. Some occurred the ***following*** year. A second example flows from our discussion of use of the mid-year population exposed to risk as an estimate of the mean population exposed to risk, and hence an estimate of person-years exposed to risk in the calculation of a crude death rate. If a person dies prior to 1 July in the reference year, her/his death is recorded in the numerator of the CDR, but (s)he is not alive at mid-year to be part of the exposed-to-risk denominator.

The first of these examples concerns an estimate of a probability; the second concerns a rate. Rates (sometimes called ***central*** rates), when measuring mortality, invariably breach the principle of correspondence in the manner described. Probabilities, when constructed without using the type of estimation procedure the IMR uses, do not. They are inherently cohort measures, which focus on a group of individuals and demographic events experienced by that group. The distinction being made here should become clearer once you are familiar with Lexis diagrams.

Other Types of Measures

We have focused at some length on demographic rates and probabilities because of their central importance in demographic analysis. However, discounting absolute measures, also already discussed, there is a range of other types of measures that interest demographers. Many are, like rates and probabilities, ratios. One that is commonly encountered is the sex ratio, giving the number of males per 100 females in a population or population subgroup:

$$S = (M/F)\ .\ 100 \tag{1.14}$$

Where $M =$ number of males in the population or subgroup; $F =$ number of females.

While rates and probabilities have numerators which are counts of ***events*** and denominators which are counts of ***people*** at risk (or person-years of exposure to risk), both numerator and denominator of the sex ratio are counts of people. Several other measures fit this pattern; for example, the dependency ratio (the ratio of persons in the dependent age groups per 100 in the working age groups), and the child-woman ratio (a fertility measure sometimes used if counts of birth events are not available). These are calculated as follows:

$$D = ((E + Y)/W)\ .\ 100 \tag{1.15}$$

Where $E =$ persons in the age group defined to constitute the dependent elderly (often ages 65 and over); $Y =$ persons in the age group defined to constitute the dependent young (often ages 0–14); $W =$ persons in the 'working' age group (often ages 15–64).

The dependency ratio can also be split into additive components, the old-age dependency ratio:

$$D_{(\text{old-age})} = (E/W)\ .\ 100 \tag{1.16}$$

and the youth dependency ratio:

$$D_{(\text{youth})} = (Y/W)\ .\ 100 \tag{1.17}$$

The child-woman ratio, the number of children aged 0–4 per 1,000 women of childbearing age, is given by:

$$CWR = \left(P_{(0-4)}/P_{(f,15-49)}\right)\ .\ 1{,}000 \tag{1.18}$$

Where $P_{(0\text{–}4)} =$ population (of both sexes) aged 0–4; $P_{(f,15\text{–}49)} =$ female population aged 15–49.

Other measures are not single ratios of any description, but ***sums*** of series of ratios. The best known of these is the total fertility rate (TFR). In recent years this widely used index has been renamed in some quarters the total fertility ***ratio***, in recognition that it does not conform to the strict definition of a demographic rate. But even this name misrepresents it; as just indicated it is ***not*** a single ratio, but the sum of ***several*** ratios. The TFR is the sum, over all reproductive ages, of the age-specific fertility rates prevailing in a particular calendar year; that is:

$$\mathrm{TFR} = \Sigma_{(\mathrm{x}=15,49)}\ \mathrm{ASFR_x} \tag{1.19}$$

Where $\mathrm{ASFR_x}$ = age-specific fertility rate for women aged x; x = a single-year age group.

Alternatively the TFR may be calculated using age-specific fertility rates for women in five-year age groups:

$$\mathrm{TFR} = 5\Sigma_{(\mathrm{x}=15-19,45-49)}\ \mathrm{ASFR_x} \tag{1.20}$$

The TFR is, in fact, a measure of the number of children the average woman in a population would have ***throughout her life, IF*** she experienced the age-specific fertility rates that prevailed in the year for which the TFR is being calculated. The multiplier of 5 that appears in Eq. 1.20 recognizes that it takes a woman 5 years to pass through a 5-year age group, so she is therefore exposed to the ***annual*** fertility rate for that age group five times.

The TFR and similar measures (e.g., the total first marriage rate (TFMR) – the sum over all marriageable ages for a given sex of ratios first marriages at age x in a calendar year divided by the mid-year total population of that sex aged x) are summary measures of the ***level*** of a set of specific rates or ratios. Other measures summarize the ***distribution*** of sets of specific measures, an important group of such measures being what statisticians call ‘measures of central tendency’ – mean (or on occasion, median) ages or durations at which specified demographic events occur in a population. Examples of such measures are the mean age of women at first birth, the mean age of childbearing, the median age at first marriage, and the median duration of first marriage at divorce. Shifts in these indices for a population over time, or differences between indices for discrete populations at a point in time, indicate changes or differences in the ***timing*** of the demographic process in question.

This brief discussion by no means exhaustively covers ‘other’ types of measures used by demographers. It is, however, indicative of some major types, and along with the earlier discussion of rates and probabilities, is sufficient to demonstrate that demographic analysis utilizes a range of distinctive categories of measure.

References

Australian Bureau of Statistics. (2007). *Deaths Australia 2006*. Catalogue 3302.0. Canberra: Australian Bureau of Statistics.

Bogue, D. J. (1969). *Principles of demography*. New York: Wiley.

Carmichael, G. A. (1986). Birth order in Australian vital statistics. *Journal of the Australian Population Association, 3*(1), 27–39.

Carmichael, G. A. (1993). Beware the passenger card! Australian and New Zealand data on population movement between the two countries. *International Migration Review, 27*(4), 819–849.

Cleland, J. G., & Scott, C. (1987). *World fertility survey: An assessment of its contribution*. New York: Oxford University Press.

Corr, P., & Kippen, R. (2006). The case for parity and birth order statistics. *Australian and New Zealand Journal of Statistics, 48*(2), 171–200.

Festy, P., & Prioux, F. (2002). *An evaluation of the fertility and family surveys project*. New York/Geneva: United Nations.

Hauser, P. M., & Duncan, O. D. (1959). The nature of demography. In P. M. Hauser & O. D. Duncan (Eds.), *The study of population: An inventory and appraisal* (pp. 29–44). Chicago: University of Chicago Press.

Hinde, A. (1998). *Demographic methods*. London: Arnold.

International Union for the Scientific Study of Population. (1958). *Multilingual demographic dictionary: English section*. New York: United Nations Department of Economic and Social Affairs.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modelling population processes*. Oxford: Blackwell.

Shryock, H. S., Siegel, J. S., & Associates. (1973). *The methods and materials of demography*. 2nd Printing (revised). Washington, DC: U.S. Bureau of the Census.

Siegel, J. S., & Swanson, D. A. (2004). *The methods and materials of demography* (2nd ed.). San Diego/London: Elsevier Academic Press.

United Nations. (1955). *Manual II. Methods of appraisal of quality of basic data for population estimates* (Population Studies No. 23). New York: United Nations. Available online at http://www.un.org/esa/population/techcoop/DemEst/manual2/manual2.html. Accessed 26 July 2013.

United Nations. (1983). *Manual X. Indirect techniques for demographic estimation*. ST/ESA/SER.A/81. New York: United Nations. Available online at http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html. Accessed 26 July 2013.

United Nations. (2001). *Principles and recommendations for a vital statistics system: Revision 2*. New York: United Nations.

United Nations Statistics Division. (2013). *2010 World population and housing census program*. Available online at http://unstats.un.org/unsd/demographic/sources/census/censusdates.htm. Accessed 12 July 2013.

Verhoef, R., & van de Kaa, D. (1987). Population registers and population statistics. *Population Index, 53*(4), 633–642.

Weeks, J. R. (1994). *Population: An introduction to concepts and issues* (5th ed.). Belmont: Wadsworth.

Weinstein, J., & Pillai, V.K. (2001). *Demography: The science of population*. Boston: Allyn and Bacon.

Wunsch, G. J., & Termote, M. G. (1978). *Introduction to demographic analysis principles and methods*. New York: Plenum Press.

Chapter 2

Comparison: Standardization and Decomposition

Population Composition and Comparison of Summary Measures

Demographers regularly wish to make ***comparisons***, either between two or more populations at the same point in time (or thereabouts) or between the same population at two or more points in time. They frequently seek to do this by way of some single ***summary measure*** calculated for each of the populations to be compared. Comparisons of summary measures can, however, be quite misleading because differences between them reflect differences in ***population composition***, as well as differences in the ***underlying level, incidence or prevalence*** of the demographic phenomenon under scrutiny. 'Population composition' refers to the proportionate distribution of a population across categories of variables, or combinations of variables, which capture demographically pertinent characteristics of individuals. Such variables include age, sex, marital status, ethnicity, education and occupation, but not, for example, variables like eye colour, or whether left or right handed. The most familiar representation of population composition is the ***age-sex pyramid***, a graphical device for summarizing the proportionate distribution of a population by age and sex, several examples of which were presented in Chap. 1 to illustrate the capacity of census data to capture demographic change over time. Composition is, however, a multidimensional concept that embraces several other variables besides these two.

The range of 'summary measures' to which the problem addressed in this chapter might apply is extensive. It includes everything from basic measures like the crude death and birth rates through more refined measures such as the general fertility rate to even more refined measures such as, for example, the age-specific fertility rate for women aged 15–19. Note that the word 'rate' recurs in these examples. One of them, however, is not a 'true' rate (the CBR), and there are other measures which are neither called, nor satisfy our definition of, 'rates' to which the principles

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_2

and techniques about to be discussed also apply. Examples are the proportion of women of marriageable age who have never married, and the mean number of children ever born to women of childbearing age or older. So what ***does*** distinguish the 'summary measures' that are at risk of yielding distorted comparisons, and hence are of interest? Essentially they are ***ratio-type measures that have rate-type denominators***. That is, their denominators are, or can be approximated as, the size of a population, or a population subgroup, ***at a point in time***. They thus include all genuine 'rates' (whose denominators typically are approximated as the population at risk at the middle of the reference period – a point in time), plus a range of other measures with similar types of denominators (but numerators which are not counts of events during a reference period). ***Excluded***, however, ***are probabilistic types of measures***, whose denominators are not populations at a point in time, but populations attaining an exact age or duration over a ***period*** of time. Discussion of Lexis diagrams in Chap. 3 should make this distinction clearer.

The values summary measures of the type just described take on are the product of two things: (i) the underlying level, or intensity, of the demographic phenomenon being measured; and (ii) the composition of the (denominator) population for which a calculation is made (i.e., the extent to which that population is concentrated in compositional categories where the phenomenon being measured is especially likely or unlikely to occur). In comparing populations we generally are interested in how ***underlying*** levels or intensities compare, but we have to accept that summary measures for those populations normally differ only ***partly*** as a function of differences therein; they also differ as a function of differences in population composition. Obvious questions arise: What are the relative strengths of the two sources of difference? Are they operating in the same direction, or is the observed difference the net effect of opposing forces, each favouring a different population? If the latter, which of the two forces has proved stronger and therefore dictated the direction of the observed difference?

The claim that summary measures of the type under discussion have an 'underlying level, or intensity' component and a 'composition' component can be substantiated by a simple example. Consider the crude death rate, the equation for which was given in Chap. 1 (Eq. 1.6) as CDR = (D / P). 1,000, where D = deaths during year y and P = the mid-year total population in year y. Death is a phenomenon the incidence of which is known to vary markedly by age and, to a lesser extent, sex (two dimensions of population composition). We can rewrite the equation for the CDR as follows (dropping the multiplier of 1,000):

$$\mathrm{CDR} = \Sigma_{\mathrm{s}} \Sigma_{\mathrm{x}} \ (\mathrm{d}\,(\mathrm{s},\mathrm{x}) \,/\mathrm{P}) \tag{2.1}$$

Where s denotes sex (male or female); x denotes an age group; d(s,x) = deaths of persons of sex s at age x during year y.

All we have done in Eq. 2.1 is say that total deaths in year y (i.e., D) equals the sum of deaths in each sex-age group during year y (i.e., $\Sigma_{\mathrm{s}} \Sigma_{\mathrm{x}}$ d(s,x)). Effectively we divide the deaths in each sex-age group by P then add up the answers, instead of adding the deaths first and then dividing the total by P; the answer is the same.

We now perform a little trick and multiply the righthand side of Eq. 2.1 by a quantity (p(s,x) / p(s,x)) (where p(s,x) is the mid-year population of sex s aged x in year y). We can do this because plainly this quantity has the value 1 (its numerator and denominator are identical and therefore cancel), so that the value of the righthand side is unaltered. We get:

$$\mathrm{CDR} = \Sigma_{\mathrm{s}}\Sigma_{\mathrm{x}}\ \{(\mathrm{d}\,(\mathrm{s},\mathrm{x})\,/\mathrm{P})\,(\mathrm{p}\,(\mathrm{s},\mathrm{x})\,/\mathrm{p}\,(\mathrm{s},\mathrm{x}))\}$$

Which can be rearranged to give:

$$\mathrm{CDR} = \Sigma_{\mathrm{s}}\Sigma_{\mathrm{x}}\ \{(\mathrm{d}\,(\mathrm{s},\mathrm{x})\,/\mathrm{p}\,(\mathrm{s},\mathrm{x}))\,(\mathrm{p}\,(\mathrm{s},\mathrm{x})\,/\mathrm{P})\} \tag{2.2}$$

The righthand side of Eq. 2.2 now incorporates the product of two ratios, d(s,x) / p(s,x), which is the sex-age-specific death rate for sex s and age group x, and p(s,x) / P, which gives the proportion of the total mid-year population that has sex s and age group x. Thus, we have derived an equation for the CDR in terms of ***specific*** death rates (death rates specific for categories defined by two compositional variables, sex and age, across which we expect the incidence of death to vary) and ***population composition*** (here, sex-age composition).

It is possible to do exactly the same for any other summary measure whose denominator is the size of a population, or a population subgroup, at a point in time. With the CDR we have decided that age and sex are key compositional dimensions across which variation in the level of mortality is likely, and have consequently expressed it as the ***sum*** of the ***products*** of (i) death rates specific for categories defined by those two variables, and (ii) proportions of total population in the same categories. Similarly, we can express ***any other*** summary measure with the appropriate type of denominator as the sum of the products of two things:

1. Equivalent ***specific measures*** for categories which are defined by all possible combinations of the categories of individual compositional variables held to be important. The criterion for selecting a compositional variable as important is an expectation that the intensity of the process or status in question (mortality, fertility, education, etc.) varies markedly across its categories. This series of measures collectively captures the ‘underlying level, or intensity’ component of the summary measure referred to above.
2. Measures of ***population composition***; i.e., proportions of the population forming the denominator of the summary measure in each of the categories for which specific measures are calculated. Obviously these proportions should sum to 1.0.

This specification gives rise to the following generalized equation:

$$\mathrm{M} = \Sigma_{\mathrm{c}}\ \mathrm{m}\,(\mathrm{c})\,.\mathrm{p}\,(\mathrm{c}) \tag{2.3}$$

Where M denotes a summary measure with the appropriate type of denominator; c denotes a series of compositional categories defined as specified under point 1 above; m(c) = the specific measure equivalent to M for compositional category c; p(c) = the proportion of the population forming the denominator of M who are members of compositional category c.

Before moving on, some discussion of compositional variables and categories is in order. ***The compositional variable demographers most often suspect of affecting the values of summary measures, and thereby contaminating comparisons of those measures, is age***. Most demographic processes are stronger, and demographic statuses more common, at some ages than at others, and it is routine to control comparisons for differences in age composition. ***Do not***, however, fall into the trap of thinking that age is the ***only*** compositional variable you ever need to worry about. In dealing with the CDR above we also took into account the variable sex, and there is ample evidence that mortality levels also vary across compositional dimensions such as ethnicity, education, occupation and marital status. We might, in certain circumstances, wish to recognize one or more of these additional variables in specifying the CDR in accordance with Eq. 2.3 (so, for example, our categories, (c), might be defined in terms of age, sex ***and ethnicity***).

Or take a summary measure like the age-specific fertility rate for females aged 15–19. Here a ‘specific’ measure might also be treated as a ‘summary’ measure, because there is reason to suspect significant variation in the level of fertility among subgroups of the female population aged 15–19. The compositional variable we might be most interested in is marital status, since we expect married women in this age group to more often give birth than unmarried women. Thus, in specifying the equation for this age-specific fertility rate in accordance with Eq. 2.3 our compositional categories (c) would be marital status categories. Age, though, might also be a consideration. Within the age range 15–19 years, women are more likely to have children if aged 19 than if aged 15. If we suspected populations of 15–19 year-old females we wished to compare of having different ***single***-year age distributions, we might want to recognize both marital status ***and*** age in specifying the age-specific fertility rate in accordance with Eq. 2.3. The compositional categories would then be age-marital status categories.

Implicit in the previous two paragraphs is the fact that, in specifying a summary measure in the form generalized in Eq. 2.3 (as a precursor to applying techniques which allow one to ***control*** for the distorting (or ***confounding***) influence of differences in population composition on comparisons of summary measures), ***the decision as to what variables to take into account in defining the compositional categories (c) is arbitrary***. Often, in practice, only one variable is deemed to be of concern, most commonly age. If larger numbers of variables are thought to be potentially important, a useful strategy is to take them into account ***serially*** (exploring the impact of controlling for one variable, reworking the analysis to also take a second variable into account, etc.), and in ***different combinations***. By a process of trial and error one can isolate, and rank, the variables that are really important confounders. But ***it rarely makes sense to take large numbers of***

compositional variables into account simultaneously, especially as a first step. A common practical difficulty is obtaining the detailed cross-tabulated data needed for analysis using multi-dimensional compositional categories. More fundamentally, though, to conclude that having controlled for a 'grab-bag' of compositional variables alters the conclusion drawn ***before*** controlling is not all that helpful if one doesn't know ***which*** variable(s) is/are primarily responsible for the finding.

It should also be appreciated that controlling a comparison for one or more compositional variables using methods with Eq. 2.3 at their core does not guarantee that distortion due to differences in population composition has been eliminated. First, ***other compositional variables not taken into account may remain sources of distortion***. Sometimes, lack of data precludes controlling for an important variable; on other occasions an analyst may fail to recognize the need to control for a variable. Second, taking a compositional variable into account using Eq. 2.3 isn't just a matter of introducing it to the equation. ***It is also a matter of deciding what categories to specify for the variable***. Sometimes these are self-evident. With the variable sex, for example, there isn't much alternative to 'male' and 'female'. But for a variable like age, options exist. If calculations are to be performed manually, there is an incentive to limit the number of categories, and hence the quantity of mechanical work. However, if too few, rather broad, categories are designated, ***within which*** there is significant variation in both the level/intensity of the demographic phenomenon being studied and the distribution of individuals from population to population, the variable could, in effect, be only ***partially*** controlled for. Thus, for example, using 5-year age groups to capture age composition provides a more robust control for age than using, say, age groups 0–49 and 50 and over.

One final point relevant to selecting appropriate compositional variables should be mentioned. The distorting impact of a compositional variable on comparisons of a summary measure is a function of two things: (i) variation in the level/intensity of the demographic phenomenon under scrutiny across its categories; and (ii) variation in the distribution of members of the populations being compared among those categories. If, for a given variable, only ***one*** of these conditions exists, ***there will be no distorting impact of that variable on a comparison***. In particular, should the first condition exist without the second, the variable can be ignored. Take the example of the age-specific fertility rate for women aged 15–19 discussed above. We noted that fertility was likely to vary significantly between 15 year-olds and 19 year-olds. But for this variation to be a problem when comparing fertility rates for 15–19 year-olds there needs to also be significant variation ***from population to population*** in the proportionate distribution of 15–19 year-olds among the five constituent single-year age groups. If investigation showed these age distributions to be similar, the case for controlling for age would dissipate.

Making More Meaningful Comparisons of Summary Measures

Having recognized the potential for comparisons of some demographic summary measures to lead us astray, we turn now to techniques for dealing with the problem of distortion arising from differences in population composition. ***Three approaches*** are covered: the calculation of a series of specific measures; standardization; and decomposition (sometimes also referred to as ‘component analysis’ or ‘components of difference analysis’).

The first of these approaches was in essence covered in Chap. 1. In the section on ‘Refinement of Demographic Rates and the Effect of Age Structure’ (pp. 31–36), an example was introduced in which crude death rates for Malaysia and Australia in 1988 were compared. The CDR for Australia was higher, which seemed strange given Australia’s status as a ‘developed’ country, but when age-sex-specific death rates were examined (Table 1.2), with few exceptions they were ***lower*** for Australia than for Malaysia. It was pretty clear that, age for age and for both sexes, mortality was lower, not higher, in Australia (i.e., the ***underlying*** level of mortality was lower). We then investigated ***how*** lower age-sex-specific death rates in Australia could produce a larger number of deaths among each 1,000,000 Australians than among each 1,000,000 Malaysians, thus resulting in a higher CDR.

The Malaysian-Australian CDRs example is a spectacular one. Because of this, the move from comparing CDRs to comparing age-sex-specific death rates is enlightening. It tells us that the CDRs are totally misleading, to the point of suggesting that the mortality differential between the two countries runs in the opposite direction to the one it actually runs in. It doesn’t, however, give us any real sense of ***how much*** lower the underlying level of mortality is in Australia (or how much higher it is in Malaysia). We could easily generate a ***series*** of such measures for each age-sex group (by taking ratios of age-sex-specific death rates for the two countries), but we have no ***single*** summary indicator of the relative underlying levels of mortality.

Viewed more generally, the practice of calculating series of specific measures to assess the accuracy of the picture gained from a comparison of summary measures is also beset by other difficulties. First, with the Malaysian-Australian CDRs example we were fortunate that most of the age-sex-specific mortality differentials flowed in one direction. If differentials in specific measures flow in opposite directions, some favouring one population and others the other, unless there is a clear majority flowing one way it can become difficult if not impossible to tell in which direction the net ***overall*** differential flows. The fact that differences of direction exist, and their pattern and magnitude, are useful pieces of supplementary information that one might wish to consider in conjunction with a ***measure*** of net overall differential (which would suppress such detail). But they do not, of themselves, ***yield*** such a measure. Second, while with the Malaysian-Australian CDRs example we find a complete reversal of the ***direction*** of the mortality differential once allowance is made for differences in population composition, a more common scenario sees no change to the direction of a differential, but some modification of its ***magnitude***. Under this scenario, differences in population composition simply exaggerate or moderate a differential. This more subtle type of distorting influence can be difficult, and very often impossible, to detect merely through the calculation and comparison of specific measures. Differentials between specific measures for compositional categories will tend to run in the same direction as that between the summary measures, and judging whether the former are collectively more marked than, similar to, or more subdued than the latter may be well nigh impossible.

Thus, calculating specific measures is a hit-or-miss approach to assessing the distorting impact of population composition on a comparison of summary measures. Moreover, it is probably more often ‘miss’ than ‘hit’, and even when informative, lacks the ‘summary’ character sought when using summary measures.

Standardization

Standardization is a technique which enables us to take into account and control for the varying compositions of populations we wish to compare in the course of calculating a ***single*** index for each population. It thus has the dual advantage of being both a more robust approach to dealing with compositional distortion in comparisons, and offering ‘summary character’. It has two variants. The more straightforward, as its name implies, is ***direct*** standardization; the alternative approach is ***indirect*** standardization.

Both approaches have at their core the generalized equation expressing a summary measure in terms of specific measures and population composition presented above as Eq. 2.3, ***and that equation is therefore a very important one***. With direct standardization a ***standard*** (i.e., common) ***population composition*** is assumed for all the populations to be compared (i.e., the set of p(c) values in the righthand side of Eq. 2.3 remains the same from calculation to calculation), and the observed set of specific measures for compositional categories for each population is applied to this standard composition in turn. With indirect standardization a ***standard set of specific measures*** for compositional categories is assumed for all the populations to be compared (i.e., the set of m(c) values is the same in each calculation), and is applied, in turn, to the observed population composition for each population.

To put this in more concrete terms, suppose we want to use standardization to eliminate the effect of differences in ***age structure*** from a comparison of summary measures. ***To standardize directly for age*** we apply ***different*** sets of age-specific measures observed for the populations in question to a ***standard*** age structure, whereas ***to standardize indirectly for age*** we apply a ***standard*** set of age-specific measures to the observed ***different*** age structures for the populations.

Direct Standardization

As the more straightforward variant, direct standardization best enables one to understand how standardization works. With direct standardization, in place of our ***un***standardized summary measures we calculate ***standardized equivalents***; equivalents ***standardized for the compositional variable(s) we have chosen to take into account***. It is important to appreciate that these standardized summary measures are ***hypothetical*** summary measures. Their values are values we would ***expect*** the summary measure to take ***if*** the standard population composition, rather than the actual composition, prevailed.

What does this mean? It means that ***no significance can be attached to the absolute levels of standardized summary measures***. These will vary depending on the standard selected, and this selection is an arbitrary decision, although one often governed by fairly commonsense options. When we interpret standardized summary measures it is not their absolute levels but their ***relative*** levels that are important. We are interested in how one standardized measure ***compares*** with another – whether it is higher or lower, and by what order of magnitude. ***Standardization is a procedure designed solely to facilitate COMPARISON. It seeks to overcome the reality that differences in population composition can render comparisons of summary measures misleading by ELIMINATING those differences from the comparison.*** As implied earlier, this elimination may be less than total. Its completeness depends on our (i) choosing the appropriate variable(s) to standardize for and (ii) selecting suitably refined categories for those variables. But the major point to grasp here is that ***STANDARDIZED SUMMARY MEASURES HAVE NO MEANING OUTSIDE A COMPARATIVE SETTING***. Furthermore, they can only be compared ***if all measures were standardized using the same standard***. It is not appropriate to choose one standard composition for one population and a different one for the population you wish to compare it to. To do so would run counter to the aim of ***eliminating compositional difference*** from a comparison.

What data are required to be able to undertake a direct standardization? Given the summary measure to be standardized, we require:

1. For ***each*** population we wish to compare, a set of ***equivalent*** measures (rates, proportions, etc.) specific for compositional categories defined using the variable(s) we wish to standardize for.
2. Preferably, although not essentially, the composition (i.e., the proportionate distribution among compositional categories) of at least ***one*** of the populations we wish to compare. This population becomes the ***standard*** population. As intimated above, selecting the standard population is an arbitrary decision. The composition of some other population may be used, and in some circumstances is preferable (e.g., the national population when comparing several geographic or socio-demographic subgroups of a national population), but as discussed below there is advantage in choosing one of the populations to be compared.

If this is done, a set of specific measures for the population chosen as standard is not needed. There is no need to calculate a standardized summary measure for that population – it is identical to its ***un***standardized summary measure, whose value will already be known. For this standard population we would be standardizing to ***its own*** composition, which intuition tells us is a 'no change' situation. The advantage of selecting one of the populations to be compared as the standard population should now be clear – one less standardization calculation is necessary.

To illustrate, suppose we wanted to directly standardize the ***general fertility rates*** of two populations for ***age*** and ***marital status*** (recall that the GFR is live births during year y divided by the mid-year female population aged 15–49 in year y – Eq. 1.10). Suppose further that we decided that our age categories would be the seven 5-year age groups 15–19 through 45–49, and our marital status categories would be 'never married', 'currently married' and 'formerly married'. To carry out the necessary calculations we would need:

1. Fertility rates for each possible combination of our seven age and three marital status categories (i.e., 7 × 3 = 21 age-marital status-specific rates) for both of the populations to be compared (unless we had determined to make one of them the standard population – we would then not need specific rates for that population).
2. Preferably, but not essentially, the proportionate distribution of one of the populations among the same 21 categories (we ***could*** use as our standard population composition ***any*** plausible proportionate distribution among those 21 categories).

Generalizing, we will denote any summary measure for population i that we wish to standardize by M(i). The equation for calculating a directly standardized value of M(i) is:

$$M_s(i) = \Sigma_c\, m_i(c) .p_s(c) \tag{2.4}$$

Where $M_s(i)$ denotes the standardized summary measure for population i (standardized to the composition of population s); c denotes compositional categories defined using the variable(s) for which we are standardizing (they might be age categories, marital status categories, age-sex categories, age-marital status categories, etc.); $m_i(c)$ = the ***specific*** measure equivalent to M(i) for category c and population i; $p_s(c)$ = the proportion of the standard population s in category c.

It is of interest to compare the righthand side of Eq. 2.4 and the expression for M(i) (the ***un***standardized summary measure for population i) which is obtained using the generalized Eq. 2.3. Using Eq. 2.3, the value of M(i) is given by:

$$M(i) = \Sigma_c\, m_i(c) .p_i(c)$$

Where $m_i(c)$ and $p_i(c)$ are the values of m(c) and p(c) for population i.

The only difference between the righthand sides of this expression and Eq. 2.4 is that in Eq. 2.4 $p_i(c)$ has been replaced by $p_s(c)$. The element $p_i(c)$ gives the ***actual*** composition of population i. You can thus perhaps appreciate that the way direct standardization works is to ***eliminate*** differences in composition from a comparison by ***replacing*** the actual population composition element by the composition element for the standard population. This process is ***repeated*** in calculating directly standardized summary measures for ***each*** population being compared. The result is a series of (standardized) measures with ***identical*** composition elements. It follows that any differences between them are attributable to differences in the schedules of $m_i(c)$ values; that is, solely to differences in composition-specific levels of the phenomenon of interest (mortality, fertility, etc.), or solely to differences in the ***underlying level*** of that phenomenon.

You can now, perhaps, also better appreciate why there is no need to calculate a standardized summary measure for the standard population. For the standard population s, Eq. 2.3 gives the unstandardized summary measure M(s) as:

$$M(s) = \Sigma_c\, m_s(c).p_s(c)$$

The process of direct standardization entails replacing the p(c) element in the righthand side of this equation with the p(c) element for the standard population; $p_s(c)$ is replaced by $p_s(c)$, or by itself. Thus we have, applying Eq. 2.4:

$$M_s(s) = \Sigma_c\, m_s(c).p_s(c)$$

Clearly no change occurs. From the preceding expressions $M_s(s) = M(s)$ (the standardized and unstandardized summary measures for the standard population are equal), and as the value of M(s) is already known, no calculation using Eq. 2.4 to find $M_s(s)$ is necessary.

How do we use standardized summary measures to make comparisons which are uncontaminated by differences in population composition (at least insofar as we have standardized for the appropriate variables, and chosen appropriate categories for them)? A useful device is ***standardized ratios***. A standardized ratio is ***the ratio of standardized summary measures*** (based on the same standard) ***for two populations we wish to compare*** (i.e., the standardized summary measure for the first population divided by that for the second population). It can be placed alongside the equivalent ***un***standardized ratio (the ratio of the unstandardized summary measures for the two populations), interpretation then proceeding on two fronts:

1. The standardized ratio indicates the 'true' relative underlying level of the demographic phenomenon under investigation in the first population compared to the second.
2. Both ratios, looked at together, enable us to assess how standardization modifies the conclusion as to the relative levels of the phenomenon under investigation invited by the two unstandardized summary measures.

Standardized and unstandardized ratios ***give a value for the population providing the numerator assuming a value of 1.0 for the population providing the denominator***. If a ratio has a value ***less*** than 1.0, the phenomenon under investigation has a ***lower*** intensity in the numerator population; if it has a value ***greater*** than 1.0 the phenomenon has a ***higher*** intensity in the numerator population. The difference between a ratio and 1.0 gives a proportionate measure of the ***degree*** to which the phenomenon is less intense, or more intense, in the numerator population compared to the denominator population.

When examining how standardization modifies the conclusion invited by a comparison of two unstandardized summary measures, we are basically interested in how the standardized ratio differs from the unstandardized ratio ***relative to a pivotal value of 1.0***. First, is it closer to or further away from 1.0? Second, is it on the same side of 1.0 or on the opposite side? The ratio value 1.0 is 'pivotal' because that is the value that implies ***no difference*** between the populations being compared. We are interested in whether the standardized ratio (i) is closer to 1.0 than the unstandardized ratio (in which case standardization shows the two populations to be more similar than the original summary measures indicated) or (ii) is further away from 1.0 (in which case standardization shows the two populations to be more different than we had thought), and (iii) in whether the two ratios both lie above, both lie below, or lie on opposite sides of 1.0. Should the latter be the case, ***standardization has altered the DIRECTION of the differential between the populations, as well as its MAGNITUDE***.

Naturally, when calculating pairs of a standardized and an equivalent unstandardized ratio, it is essential that the same populations provide the numerator and denominator in each calculation. A common approach when carrying out this type of interpretative exercise after standardizing several summary measures using the same standard population is to ***compare each population in turn with the standard population***. We thus end up with a series of standardized ratios that are directly comparable, because they are based on a common comparator population, the standard population. In our earlier notation, such standardized ratios take the form:

$$\mathrm{SR} = \mathrm{M_s}\,(\mathrm{i})\,/\mathrm{M}\,(\mathrm{s}) \tag{2.5}$$

Despite contrary appearances, this is a ratio of two standardized summary measures because, as we have seen, M(s), the unstandardized summary measure for the standard population, is equal to $\mathrm{M_s}$(s), the standardized summary measure for that population. The equivalent unstandardized ratio is given by:

$$\mathrm{UR} = \mathrm{M}\,(\mathrm{i})\,/\mathrm{M}\,(\mathrm{s}) \tag{2.6}$$

When the summary measure being standardized is the crude death rate, and it is being standardized for age, a standardized ratio calculated using Eq. 2.5 is commonly referred to as the ***comparative mortality factor (CMF) for population i***. It is a measure of how much higher, or lower, mortality is in population i than in the standard population after standardizing for age. It can be shown to equal the ratio of the number of deaths ***expected*** in the standard population, assuming it had experienced the age-specific death rates of population i, to the ***actual*** number of deaths in the standard population. The name ‘comparative mortality factor’ captures the capacity of this measure, calculated for a series of populations (i, j, k, l, etc.), to facilitate legitimate ***comparison*** among those populations. If we substitute the letters D and d for the letters M and m in our generalized Eq. 2.4 (in recognition that we are now dealing with crude, and specific, ***death*** rates), we have:

$$\mathrm{D_s\,(i) = \Sigma_c\ d_i\,(c)\,.p_s\,(c)}$$

This can be rewritten:

$$\mathrm{D_s\,(i) = \Sigma_c\ d_i\,(c)\,.\,(P_s\,(c)\,/\,\Sigma_c\ P_s\,(c)) = \Sigma_c\ d_i\,(c)\,.P_s\,(c)\,/\,\Sigma_c\ P_s\,(c)}$$

Where $\mathrm{P_s(c)}$ (as distinct from $\mathrm{p_s(c)}$) is the observed population in category c in the standard population (as distinct from the ***proportion*** of total population in category c).

If we multiply the righthand side by $\mathrm{\Sigma_c d_s(c).\,P_s(c)/\Sigma_c d_s(c).\,P_s(c)}$ (which clearly equals 1) and then rearrange, we get:

$$\mathrm{D_s\,(i) = (\Sigma_c\ d_s\,(c)\,.P_s\,(c)\,/\,\Sigma_c\ P_s\,(c))\,.\,(\Sigma_c\ d_i\,(c)\,.P_s\,(c)\,/\,\Sigma_c\ d_s\,(c)\,.P_s\,(c))}$$

And as $\mathrm{\Sigma_c\,d_s\,(c)\,.P_s\,(c)\,/\,\Sigma_c\,P_s\,(c) = \Sigma_c\,d_s\,(c)\,.\,(P_s\,(c)\,/\,\Sigma_c\,P_s\,(c)) = \Sigma_c\,d_s\,(c)\,.\,p_s\,(c) = D\,(s)}$ (the crude death rate for the standard population, from Eq. 2.3):

$$\mathrm{D_s\,(i) = D\,(s)\,.\,(\Sigma_c\ d_i\,(c)\,.P_s\,(c)\,/\,\Sigma_c\ d_s\,(c)\,.P_s\,(c))}$$

Whence:

$$\mathrm{CMF = D_s\,(i)\,/D\,(s) = \Sigma_c\ d_i\,(c)\,.P_s\,(c)\,/\,\Sigma_c\ d_s\,(c)\,.P_s\,(c)} \tag{2.7}$$

This is an equation for obtaining the comparative mortality factor directly, and verifies the claim made above that the CMF is the ratio of the number of deaths expected in the standard population, were it to experience the age-specific death rates of population i, to the observed (actual) number of deaths in the standard population. The latter figure (which we can denote by $\mathrm{O_d(s)}$) may be known directly. If it is there is no need to evaluate the denominator of Eq. 2.7. That is:

$$\mathrm{CMF = \Sigma_c\ d_i\,(c)\,.P_s\,(c)\,/O_d\,(s)} \tag{2.8}$$

Let us now proceed with an example. We will again take the comparison of crude death rates for Malaysia and Australia in 1988 that was introduced in Chap. 1. The CDRs for the two populations were:

$$\text{Malaysia CDR} = 4.93 \text{ deaths per } 1,000 \text{ mid-year population}$$
$$\text{Australia CDR} = 7.25 \text{ deaths per } 1,000 \text{ mid-year population}$$

We will call Malaysia population 1 and Australia population 2. This is an arbitrary designation; we could just as easily allocate these numbers in reverse order. As above, since we are now dealing with a particular summary measure, the CDR, we will adopt the notation D for the CDR and d for a specific death rate, instead of the generalized notations M and m (standing for summary, and specific, ***measure***). Using this notation we can first specify the equations for the two CDRs in the form generalized in Eq. 2.3. We get:

$$D(1) = \Sigma_c \, d_1 (c) \,.p_1 (c)$$

and:

$$D(2) = \Sigma_c \, d_2 (c) \,.p_2 (c)$$

We can make these equations more specific by explicitly recognizing that the compositional variables we suspect of distorting a CDR-based comparison of mortality, and for which we want to standardize, are sex and age. Denoting these by s and x respectively, we get:

$$D(1) = \Sigma_s \Sigma_x \, d_1 (s, x) \,.p_1 (s, x) \tag{2.9}$$

and:

$$D(2) = \Sigma_s \Sigma_x \, d_2 (s, x) \,.p_2 (s, x) \tag{2.10}$$

The d_1(s,x) and d_2(s,x) values needed to evaluate Eqs. 2.9 and 2.10 are the age-sex-specific death rates for Malaysia and Australia given in Table 1.2 in Chap. 1. The required p_1(s,x) and p_2(s,x) values (proportions of total population in each age-sex category) are shown in Table 2.1. Using these data Eq. 2.9 yields a CDR of 4.84 deaths per 1,000 mid-year population for Malaysia and Eq. 2.10 yields a CDR of 7.16 deaths per 1,000 mid-year population for Australia. These figures are slightly different from those given above because of the rounding of age-sex-specific death rates in Table 1.2 to one decimal place and of proportions of total population in age-sex groups to three decimal places.

Suppose we ask the following question: 'What would the CDR for Malaysia be ***if its population had the same age-sex composition as Australia's population?***' In line with our earlier more general discussion, we obtain a directly standardized CDR for Malaysia ***using Australia as the standard population*** by ***replacing*** the compositional element in Eq. 2.9 with that for Australia from Eq. 2.10. The equation we need to evaluate is:

Table 2.1 Proportions of total population in age-sex groups for Malaysia and Australia, 1988

	Males		Females	
Age group	Malaysia	Australia	Malaysia	Australia
0	.014	.008	.013	.007
1–4	.057	.030	.054	.029
5–9	.061	.038	.058	.036
10–14	.056	.039	.054	.037
15–19	.053	.044	.051	.042
20–24	.050	.041	.049	.039
25–29	.043	.043	.045	.042
30–34	.036	.040	.038	.040
35–39	.030	.039	.031	.038
40–44	.024	.036	.023	.035
45–49	.020	.028	.020	.026
50–54	.017	.024	.017	.023
55–59	.013	.023	.014	.022
60–64	.010	.022	.011	.022
65–69	.007	.017	.008	.020
70–74	.005	.013	.005	.016
75–79	.003	.009	.004	.013
80–84	.001	.004	.002	.008
85+	.001	.002	.001	.006

Source: *United Nations Demographic Yearbook, 1990* and Australian Bureau of Statistics, *Estimated Resident Population by Sex and Age: States and Territories of Australia June 1988 and Preliminary June 1989*

$$D_2(1) = \Sigma_s \Sigma_x \, d_1(s, x) \, . \, p_2(s, x) \tag{2.11}$$

You can probably appreciate that this equation has the same form as Eq. 2.4, the general equation for a directly standardized summary measure. The summary measure is the CDR (denoted by D in place of M); the compositional categories are defined by two variables, sex and age (denoted by s and x in place of c); and the standard population is population 2 (Australia's population). We can use the data in Tables 1.2 and 2.1 to evaluate this equation. In doing this we construct the calculating table shown as Table 2.2. The first column of figures is taken from Table 1.2, the second column from Table 2.1. To obtain the third column we multiply the figures from the first two columns for each sex and age group. We then add the answers in this column to obtain $D_2(1)$, our standardized death rate (this is what the righthand side of Eq. 2.11 instructs us to do).

The answer we obtain, $D_2(1) = 10.42$, is much higher than our ***un***standardized CDR for Malaysia ($D(1) = 4.84$). It is a ***hypothetical*** CDR. It tells us that ***if*** Malaysia (population 1) had had the age-sex-specific death rates actually recorded for that country in 1988, but in combination with ***Australia's*** (population 2's) rather than its own age-sex composition, its CDR would have been not 4.84, but 10.42.

Table 2.2 Calculating table for obtaining, using Eq. 2.11, a 1988 CDR for Malaysia directly standardized to the age and sex composition of the population of Australia

Sex and age group		$d_1(s,x)$	$p_2(s,x)$	$d_1(s,x).p_2(s,x)$
Males	0	16.6	.008	.1328
	1–4	1.7	.030	.0510
	5–9	0.6	.038	.0228
	10–14	0.6	.039	.0234
	15–19	1.0	.044	.0440
	20–24	1.5	.041	.0615
	25–29	1.7	.043	.0731
	30–34	2.0	.040	.0800
	35–39	2.4	.039	.0936
	40–44	3.4	.036	.1224
	45–49	5.5	.028	.1540
	50–54	9.9	.024	.2376
	55–59	15.4	.023	.3542
	60–64	25.5	.022	.5610
	65–69	38.3	.017	.6511
	70–74	62.7	.013	.8151
	75–79	86.2	.009	.7758
	80–84	144.3	.004	.5772
	85+	173.4	.002	.3468
Females	0	12.8	.007	.0896
	1–4	1.1	.029	.0319
	5–9	0.5	.036	.0180
	10–14	0.4	.037	.0148
	15–19	0.5	.042	.0210
	20–24	0.6	.039	.0234
	25–29	0.9	.042	.0378
	30–34	1.1	.040	.0440
	35–39	1.5	.038	.0570
	40–44	2.3	.035	.0805
	45–49	3.2	.026	.0832
	50–54	5.9	.023	.1357
	55–59	9.9	.022	.2178
	60–64	17.0	.022	.3740
	65–69	27.8	.020	.5560
	70–74	46.6	.016	.7456
	75–79	69.8	.013	.9074
	80–84	119.0	.008	.9520
	85+	142.8	.006	.8568
			$\Sigma_s \Sigma_x\, d_1(s,x) \cdot p_2(s,x) = 10.4239$	

We can now compare this standardized CDR for Malaysia directly with the CDR for Australia that we began with (7.16 deaths per 1000 mid-year population). Having selected Australia as our standard population, the standardized CDR for that population is identical to its unstandardized CDR. In the notation we have been using:

$$D_2(2) = D(2)$$

We can see that the standardized CDR for Malaysia ($D_2(1) = 10.42$) is higher than that for Australia ($D_2(2) = 7.16$), a differential running in the opposite direction to that between the respective unstandardized CDRs ($D(1) = 4.84$; $D(2) = 7.16$). We need, however, to carry our interpretation of this result further. It is the exception for standardization exercises to yield results as dramatic as this, and we need some more precise indicator of the impact of standardization. We need to calculate an unstandardized ratio and the equivalent standardized ratio. The former measures the relative levels of mortality in the two populations as indicated by the two unstandardized CDRs, D(1) and D(2). We have:

$$D(1)/D(2) = 4.84/7.16 = 0.68$$

In keeping with the principles for interpreting standardized and unstandardized ratios outlined earlier, this number suggests that the level of mortality in population 1 (Malaysia) is $0.68 - 1.00 = -0.32$, or 32 % ***lower***, than in population 2 (Australia) (recall that 1.00 is the ***pivotal value*** when dealing with unstandardized and standardized ratios – the ‘no difference’ value against which we assess ratios we calculate). What if we now calculate the equivalent standardized ratio? This is given by:

$$D_2(1)/D_2(2) = 10.42/7.16 = 1.46$$

Comparing the lefthand side of this equation with Eq. 2.7, and remembering that $D_2(2) = D(2)$, you should recognize this as a comparative mortality factor (CMF) (although with only two populations being compared, one of which is the standard population, there will in this case be no other CMF with which to compare it). Performing the same simple calculation as before, the indication after standardization is that the level of mortality in Malaysia is $1.46 - 1.00 = 0.46$, or 46 % ***higher***, than in Australia. ***Thus, while the unstandardized CDRs suggest that mortality was about one-third lower in Malaysia than in Australia in 1988, after controlling for differences in age-sex composition we find that it was, in fact, almost 50 % higher.***

As already noted, this is a spectacular result. It is more common to find that the ***direction*** of the differential between two populations in levels of a demographic phenomenon is unaltered by standardization, but the ***magnitude*** of the differential is modified. But another question may have occurred to you. What if we had made Australia population 1 and Malaysia population 2, or made Malaysia, rather than

Australia, the standard population? These are alternative ways of saying the same thing. Both are arbitrary decisions. For the purpose of illustration we will retain our original designation of Malaysia as population 1 and Australia as population 2, but will see what happens when we select population 1 as the standard population in place of population 2.

Our initial question is now: ‘What would the CDR for Australia be if its population had the same age-sex composition as Malaysia’s population? We obtain a directly standardized CDR for Australia by replacing the compositional element in Eq. 2.10 with that for Malaysia from Eq. 2.9. The equation we need to evaluate is:

$$D_1(2) = \Sigma_s \Sigma_x \, d_2\,(s, x) \, . \, p_1\,(s, x) \tag{2.12}$$

As before, this is an equation with the same form as Eq. 2.4, the general equation for direct standardization; we are standardizing the CDR for population 2 using population 1 as the standard population, and the compositional variables we are endeavouring to control for are again sex (s) and age (x). Evaluating this equation using relevant data from Tables 1.2 and 2.1, and with the aid of a calculating table modelled on Table 2.2, we find that $D_1(2) = 3.19$.

In the same manner as previously, we can compare this standardized CDR for Australia directly with the CDR for Malaysia we began with because, as the standard population, Malaysia’s standardized CDR is identical to its unstandardized CDR. That is:

$$D_1(1) = D(1)$$

We can see that the standardized CDR for Australia ($D_1(2) = 3.19$) is lower than that for Malaysia ($D_1(1) = 4.84$), a differential again running in the opposite direction to that between the respective unstandardized CDRs ($D(2) = 7.16$; $D(1) = 4.84$). As before, however, to summarize our finding more precisely we need to calculate an unstandardized ratio and the equivalent standardized ratio. Normally, in doing this, we would make the standard population the denominator population. However, in this instance we will reverse that to facilitate comparison of the result obtained with Malaysia as the standard population, with that obtained with Australia as the standard population (as a result our standardized ratio will not fit the definition of a comparative mortality factor). Our unstandardized ratio is the same as before:

$$D(1)/D(2) = 4.84/7.16 = 0.68$$

Our standardized ratio is:

$$D_1(1)/D_1(2) = 4.84/3.19 = 1.52$$

Having selected the other population as the standard population, a slightly different standardized ratio is obtained for comparison with the unstandardized ratio (which naturally remains unchanged). This often happens, but the broad conclusion that whereas the unstandardized CDRs suggest that mortality was about one-third lower in Malaysia than in Australia in 1988, once differences in age-sex composition are controlled for it was, in fact, about 50 % higher, remains valid. Although we have chosen a ***different*** standard, standardized the CDR for a ***different*** population, and calculated a ***different*** number, ***our general conclusion is the same***.

The obtaining of different standardized ratios does, though, warrant comment. It implies that the result of a direct standardization – that is, the ***comparative*** conclusion reached (as distinct from the numerical value of a standardized summary measure) – is partly dependent on the choice of standard population. ***What one ends up with is an indication of how two (or a series of) populations compare when A PARTICULAR common population composition is assumed for them.*** The stability of the conclusions one draws when different standard compositions are used depends on the degree of variability from compositional category to compositional category in the extent of inter-population differentials in specific measures. If, for example, age-specific death rates in population A are at all ages higher, or lower, than those in population B by about the same percentage factor, the choice of a standard age composition will have little bearing on conclusions drawn from a direct standardization of crude death rates for age. If, however, there is, say, a much wider disparity in death rates at younger ages than at middle and older ages, choice of a very young standard age structure will weight that disparity more heavily in the calculation of age-standardized death rates than will choice of a much older one, and will yield a wider standardized differential.

It follows that while the selection of a standard composition is an arbitrary decision, some options are more sensible than others. The value of choosing one of the populations being compared to provide the standard has already been demonstrated. It obviates the need to perform standardization calculations for that population. Another common approach, when comparing geographic or socio-demographic subgroups of a national population, is to use the national population as the standard population. This effectively uses the ***average*** composition of the populations being compared as the standard, and that principle can also be extended to comparisons of other populations which are not sugroups of a national population. Another lesson from the finding that choosing a different standard population can lead to different standardized ratios is to be careful in interpretation not to make too much of ***minor*** differences between an unstandardized ratio and a corresponding standardized ratio.

Reverting to our example, another question you might be asking concerns which populations formed the numerators and denominators in the unstandardized and standardized ratios we calculated. Population 1 consistently provided the numerator; population 2 the denominator. This was because (i) population 2 was initially the standard population and (ii) when we later made population 1 the standard we needed to maintain the same relativity to be able to compare outcomes. The initial designation of population 2 as the denominator population flowed from the principle that where multiple pairwise comparisons based on the same standard are to be made, using the standard population as denominator provides a common comparator, thus enabling a series of standardized ratios to be legitimately compared. This is the principle that underpins the comparative mortality factor (CMF), and the same idea can be extended to other demographic phenomena. But what if we had started with population 1 as the standard population? We'd have got the following unstandardized ratio:

$$D(2)/D(1) = 7.16/4.84 = 1.48$$

and the following standardized ratios:

$$D_1(2)/D_1(1) = 3.19/4.84 = 0.66$$
$$D_2(2)/D_2(1) = 7.16/10.42 = 0.69$$

We would have concluded that the unstandardized ratio suggested that the level of mortality was $1.48 - 1.00 = 0.48$, or 48 % higher in population 2 (Australia) than in population 1 (Malaysia), whereas the two standardized ratios indicated, on the contrary, that it was $0.69 - 1.00 = -0.31$ and $0.66 - 1.00 = -0.34$, or 31 and 34 % lower respectively. Thus, generalizing, we might say that while our unstandardized CDRs pointed to mortality being about 50 % higher in Australia, our standardized CDRs showed it to be around one-third lower. ***This is the same general conclusion as we reached before.*** Then, we treated Australia as the 'base' population and noted how much higher or lower mortality in Malaysia was, whereas this time the roles of the countries have been reversed. But if you can convince yourself that if A is 50 % higher than B, then B is one-third lower than A, you should be able to see that the two broad conclusions amount to the same thing.

Before moving on, let us briefly consider the case of comparing, not two, but several populations (1, 2, 3, …, n). An example might be a comparison of regional CDRs within a country. In this case it would make sense to treat the national population (i.e., the aggregate of all the regional populations) as the standard population. If we denote this national standard population by N, we would calculate directly standardized CDRs for the n regions as follows:

$$\begin{aligned} D_N(1) &= \Sigma_s \Sigma_x\, d_1(s,x) \,.\, p_N(s,x) \\ D_N(2) &= \Sigma_s \Sigma_x\, d_2(s,x) \,.\, p_N(s,x) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ D_N(n) &= \Sigma_s \Sigma_x\, d_n(s,x) \,.\, p_N(s,x) \end{aligned}$$

When seeking to compare regions before and after standardization the sensible approach would be to adopt the national population as a common comparator population in pairwise comparisons with all regions. Its unstandardized CDR (D(N)) would equal its standardized CDR ($D_N(N)$), and interpretation would be based on comparing a series of unstandardized ratios D(1) / D(N), D(2) / D(N), …, D(n) / D(N) with a corresponding series of standardized ratios (comparative mortality factors) $D_N(1)/D_N(N)$, $D_N(2)/D_N(N)$, …, $D_N(n)/D_N(N)$.

Indirect Standardization

Indirect standardization is conceptually more complicated than direct standardization, but in no way an inferior technique. Indeed it has an extremely appealing attribute in that, unlike direct standardization, it ***does not require specific measures to be known for each of the populations to be compared***. This is obviously an attraction when, as is not infrequently the case, such measures are simply not available. But it is also an attraction when, because of small numbers of events, cases or persons at risk in compositional categories, specific measures, although able to be calculated, are highly unstable and subject to significant random error. This quite often happens, for example, when dealing with mortality, especially when we wish to compare small population subgroups, such as, for instance, Australia's myriad birthplace groups. By the time annual deaths by birthplace are broken down by 5-year age groups and sex, numbers can get very small and subject to appreciable annual fluctuation. Age-sex-specific death rates consequently can become quite unstable, and this instability can translate into unstable direct standardization results. ***The reliance of indirect standardization on a single, standard set of specific measures***, which are likely to pertain to a much larger population (e.g., a national population), and therefore to be much more stable, is a very valuable attribute.

We will again begin by noting the data requirements in order to be able to apply this method. Given the summary measure to be standardized, we require:

1. A single set of equivalent measures (rates, proportions, etc.) specific for compositional categories defined using the variable(s) we wish to standardize for. This set of specific measures ideally will pertain to one of the populations we wish to compare, or to some aggregate of those populations (e.g., the national population when a series of geographic or socio-demographic subgroup populations are being compared). These specific measures are our ***standard*** specific measures.
2. The compositions (i.e., proportionate distributions among compositional categories) of ***each*** of the populations we wish to compare. However, if one of these populations is the source of our standard set of specific measures, we do not need the composition of that population.
3. The ***un***standardized summary measures for ***each*** of the populations being compared.

You will note that the above specification ***makes no reference to a 'standard population'***; rather it refers to the population that provides ***a standard set of specific measures***. It is a common misconception that this population in indirect standardization is the population to which the standardization is performed. With indirect standardization ***there is strictly speaking NO single standard population***. We are still calculating values that eliminate differences in population composition from a comparison, but not in a way that simply substitutes for the composition of each comparator population that of a standard population. The procedure we use is ***indirect*** (hence its name), and we do not calculate standardized summary measures comparable to those calculated in direct standardization. Rather, ***we jump***

directly to the calculation of standardized ratios, in which, it can be shown, direct standardization is implicit. But ***it is incorrect to refer to the population providing the standard set of specific measures as the 'standard population'***.

The crux of indirect standardization is the generation, for a population, of an expected value for the summary measure of interest, or an expected value for its numerator, on the assumption that the standard set of specific measures (rather than the actual set, which may be unknown) ***applied to the population's distribution among compositional categories.*** The ratio of the ***observed*** value of the summary measure (or its numerator) to this ***expected*** value is then taken, and ***is a standardized ratio*** indicating how much higher or lower the intensity of the demographic phenomenon in question is in the population of interest compared to the population providing the standard. The reason we can deal with ratios of observed to expected values of either the summary measure, or just its numerator, is that both observed and expected summary measures have the same denominator. Hence, either way, the standardized ratio has the same value, because the common denominators, if present, cancel one another out.

In general terms, an indirectly standardized ratio (ISR) for population i is given by:

$$\mathrm{ISR} = \mathrm{M}\,(\mathrm{i}) \,/\, \Sigma_{\mathrm{c}}\ \mathrm{m}_{\mathrm{s}}\,(\mathrm{c})\ .\ \mathrm{p}_{\mathrm{i}}\,(\mathrm{c}) \tag{2.13}$$

That is, ISR is the ratio of the observed summary measure for population i (M(i)) to its expected value were the standard set of specific measures ($m_s(c)$), rather than the actual set ($m_i(c)$), to apply to its observed population composition ($p_i(c)$). This equation can be rewritten as follows:

$$\mathrm{ISR} = \Sigma_{\mathrm{c}}\ \mathrm{m}_{\mathrm{i}}\,(\mathrm{c})\ .\ \mathrm{p}_{\mathrm{i}}\,(\mathrm{c}) \,/\, \Sigma_{\mathrm{c}}\ \mathrm{m}_{\mathrm{s}}\,(\mathrm{c})\ .\ \mathrm{p}_{\mathrm{i}}\,(\mathrm{c})$$

In this form we can see clearly that we have the ratio of two ***directly*** standardized summary measures – the specific measures elements (the m elements) in the numerator and denominator differ, but the population composition elements (the p elements) are the same. ***But***, here the common composition element is ***not*** that of the population providing the standard. It is that of population i. That is why the method is called ***in***direct standardization, and why it is incorrect to call the population providing the standard set of specific measures 'the standard population'. With indirect standardization we are once again seeking to make comparisons in which differences in population composition have been eliminated. But the common population composition ends up being that of population i, not that of population s. ***The 's' now means something different*** – it means the population providing a standard set of specific measures, not the population providing a standard composition. We will return to this issue below.

If we now multiply numerator and denominator by $\Sigma_c\ P_i(c)$ (taking into account that $p_i\,(c) = P_i\,(c) \,/\, \Sigma_c\ P_i\,(c)$, where $p_i(c)$ is the ***proportion*** of population i in compositional category c and $P_i(c)$ is the ***number*** of persons from population i in that category), we get:

$$\text{ISR} = \Sigma_c\ m_i(c)\ .\ P_i(c) / \Sigma_c\ m_s(c)\ .\ P_i(c)$$

Whence:

$$\text{ISR} = O(i) / \Sigma_c\ m_s(c)\ .\ P_i(c) \tag{2.14}$$

Where O(i) is the observed number of events, or persons with the demographic status of interest, in population i.

Equation 2.14 is a calculating equation for the ISR that yields the ratio of observed to expected ***numbers*** of events or persons of a particular demographic status, whereas Eq. 2.13 yields the ratio of observed to expected ***values of the summary measure***. Either equation may be used, the choice perhaps depending on the nature of available data.

The above is a ***generalized*** outline of indirect standardization. Most demographic analysis texts deal with the technique exclusively in the context of the analysis of ***mortality***, and specifically the ***standardization of crude death rates for age***. They do not present equations for the ISR, but for the ***standardized mortality ratio*** (or ***SMR***). Epidemiologists also use the acronym SMR to mean standardized ***morbidity*** ratio, when the measures they are standardizing pertain to levels of disease or illness, rather than mortality. Equations 2.13 and 2.14 can readily be rewritten to yield the standardized mortality ratio, which is merely a specific example (albeit a widely encountered one) of an ISR. In line with earlier notation, this rewriting entails substituting the letters D and d (standing for crude, and specific, ***death*** rates) for the letters M and m (standing for summary, and specific, ***measures***). Thus, we have:

$$\text{SMR} = D(i) / \Sigma_c\ d_s(c)\ .\ p_i(c) \tag{2.15}$$

And:

$$\text{SMR} = O_d(i) / \Sigma_c\ d_s(c)\ .\ P_i(c) \tag{2.16}$$

Where $O_d(i)$ is the observed number of deaths in population i.

It is usual to calculate SMRs separately for each sex, so that the CDRs standardized are sex-specific CDRs and the compositional categories, c, are age groups. It is, though, also possible to standardize conventional CDRs, in which case the categories, c, may be either age-sex groups or simply age groups.

We need now to return to the point made earlier that the calculation of an ISR, and by implication a SMR, is equivalent to standardizing ***directly***, but with population i, not population s, providing the standard (common) composition (perhaps age, or age-sex, structure). This means that if we calculate ISRs or SMRs for a ***series*** of populations (i, j, k, l, etc.), although we may use a common standard set of specific measures, ***each calculation uses a different standard population composition*** – the compositions of populations i, j, k, l, etc. respectively. We noted earlier, in discussing the results of our direct standardizations of the CDRs for Malaysia and Australia using first one, then the other, population's composition as the standard, that the values of standardized ratios don't necessarily remain constant when different standards are adopted. It follows that ***while an ISR or a SMR provides a legitimate PAIRWISE comparison between population i and population s*** (the population providing the standard set of specific measures) ***it is NOT, strictly speaking, valid to compare ISRs or SMRs for a SERIES of populations i, j, k, l, etc., even though they are calculated using the same population s.***

That said, the reality is that such comparisons commonly ***are*** made. Thus, the practical recommendation is that you should make them mindful of the technical invalidity of doing so. The advantages of indirect standardization – the fact that it can be used when specific measures for the populations of interest are either unknown or based on such small numbers of cases as to be likely to introduce serious error into a direct standardization – have to be weighed against the downside of the technical invalidity of comparing ISRs or SMRs. One piece of advice is to ***not place too much emphasis in interpretation upon relatively small differences between ISRs or SMRs***. Another is to ***be particularly careful in situations where the compositions of the populations for which ISRs or SMRs are being compared are known, or are likely, to be very different***. A situation in point might be a comparison of SMRs for Australian birthplace groups. Because migration tends to take place in early adulthood, and different birthplace groups have largely come to Australia over different periods of time, the age structures of different groups tend to be quite different. This is the sort of environment in which considerable bias deriving from the effective use of different standard age structures ***could*** enter a comparison of SMRs.

We will again finish with an example, once again comparing the 1988 CDRs for Malaysia and Australia. We will calculate a SMR in which we will again seek to control for differences in age-sex composition, and will continue to regard Malaysia as population 1 and Australia as population 2. Suppose further that we decide to treat the age-sex-specific death rates for population 2 (Australia) as the standard set of specific measures, so that we will calculate a SMR for Malaysia, comparing its mortality to that of Australia.

Using Eq. 2.15 we have:

$$\mathrm{SMR} = \mathrm{D(1)}/\,\Sigma_c\ \mathrm{d}_2\,(\mathrm{c})\ \ .\ \mathrm{p}_1\,(\mathrm{c}) = \mathrm{D(1)}/\,\Sigma_s\Sigma_x\ \mathrm{d}_2\,(\mathrm{s,x})\ \ .\ \mathrm{p}_1\,(\mathrm{s,x})$$

The denominator of this expression is the righthand side of Eq. 2.12, which was earlier evaluated using data from Tables 1.2 and 2.1 as having a value of 3.19. D(1) = 4.84, so:

$$\mathrm{SMR} = 4.84/3.19 = 1.52$$

This is identical to the standardized ratio we obtained earlier using direct standardization with Malaysia as the standard population. If, in calculating our SMR, we reverse things and make the age-sex-specific death rates for population 1 (Malaysia) the standard set of specific measures, we have:

$$\mathrm{SMR} = \mathrm{D}(2)/\Sigma_c\ \mathrm{d}_1(\mathrm{c})\ .\ \mathrm{p}_2(\mathrm{c}) = \mathrm{D}(2)/\Sigma_s\Sigma_x\ \mathrm{d}_1(\mathrm{s},\mathrm{x})\ .\ \mathrm{p}_2(\mathrm{s},\mathrm{x})$$

The denominator of this expression is the righthand side of Eq. 2.11, which was earlier evaluated using data from Tables 1.2 and 2.1 as having a value of 10.42. D(2) = 7.16, so:

$$\mathrm{SMR} = 7.16/10.42 = 0.69$$

This is the inverse of the standardized ratio (1.46) we obtained using direct standardization with Australia as the standard population. It is the inverse because indirect standardization always makes the population providing the standard set of specific measures the comparator (denominator) population. In essence, though, we have obtained the same results using indirect standardization as we did using direct standardization. Our first SMR tells us that mortality was about 50 % higher in Malaysia; our second tells us that it was about one-third lower in Australia. The two results mean the same, and as with direct standardization, each SMR could be compared with the equivalent ***un***standardized mortality ratio to assess how standardization changed the picture presented by the unstandardized CDRs.

Some Additional Points About Standardization

The following are some additional theoretical points about the standardization techniques we have discussed:

1. The categories defined for any compositional variable by which one is standardizing ***need to cover the entire population forming the denominator of the summary measure being standardized***. Take, for example, the crude birth rate (CBR), which has as its denominator the mean (or mid-year) ***total*** population. If we were standardizing the CBR for age and sex, many age-sex-specific fertility rates ($m_i(c)$ values) would be zero – those for males and for females younger than 15 and 50 or older. We would ***not***, however, be justified in using population composition ($p_i(c)$) values which expressed females in childbearing age groups as proportions only of total ***females of childbearing age***. Because the CBR uses total population as its denominator, females in childbearing age groups must be expressed as proportions of ***total*** population for standardization purposes. It would, however, be a different matter were we standardizing ***the general fertility rate*** (GFR). Then, because the denominator consists only of women of childbearing age, it is appropriate that population composition ($p_i(c)$) values should express females in childbearing age groups as proportions of total females of childbearing age.
2. Some important theoretical questions arise in standardization when a population composition variable is itself related to the demographic process to which the

summary measure being standardized relates. Direct standardization of the CDR is a good example. The age-sex composition of a population can be considered to be dependent on its level of mortality (and also its level of fertility) in the past. If change in the mortality level usually precedes change in the population (age-sex) composition, it may not really be reasonable to ask what the CDR in a country would be if its population's age-sex composition changed to that of the population of another country. Consider the questions asked earlier as the basis for direct standardization of the CDRs for Malaysia and Australia: 'What would the CDR for Malaysia be if its population had the same age-sex composition as Australia's population?' (Answer: $D_2(1)$); 'What would the CDR for Australia be if its population had the same age-sex composition as Malaysia's population?' (Answer: $D_1(2)$). In both cases we are effectively assuming that population composition can be changed without prior change in age-sex-specific mortality rates. Directly standardized summary measures are almost always interpreted as answering questions of the type we posed, but we could also interpret $D_2(1)$ to be the CDR Australia would have if Malaysia's age-sex-specific death rates prevailed, and $D_1(2)$ to be the CDR Malaysia would have if Australia's age-sex-specific death rates prevailed. These are answers to different questions. The issue is: What changes first, the death rates or population composition? If they change together, in conjunction, say, with change in fertility levels, there may not really be a reasonable question to answer using a standardized rate.

Decomposition

As already noted, decomposition is also called in the literature ***component analysis*** and ***components of difference analysis***. It is a variation of standardization and is concerned with the same general problem – making comparisons between summary measures more meaningful. While with standardization we endeavour to ***eliminate*** from a comparison the influence of one or more compositional variables, so that the comparison becomes, as far as is practicable, uncontaminated by compositional differences, with decomposition we aim to ***break down*** (decompose) a difference between two summary measures ***into components*** attributable to: (i) differences in population composition, and (ii) differences in measures specific for compositional categories. Having done this, the strategy is to assess the ***relative importance*** of the differences in composition and in category-specific measures in producing the overall difference observed.

This general description may lead you to think that we only ever decompose the difference between two summary measures into two components. That is not the case; decompositions into ***more*** than two components are not uncommon, but that is a complication we will put to one side for now. Possibly the most frequent application of decomposition is in situations where the difference between summary measures that is of interest entails ***change*** in a measure ***over time*** for a single population. In this application, change is decomposed into change due to changes in population composition and ‘real’ change (in specific measures).

It can be demonstrated quite simply that the concept of separation into components is implicit in direct standardization. Suppose we have two summary measures M(1) and M(2) which relate to populations P(1) and P(2). If we adopt P(1) as the standard population and calculate a directly standardized value of M(2) (i.e. $M_1(2)$), then it is clear that:

$$M(1) - M(2) = (M(1) - M_1(2)) + (M_1(2) - M(2)) \tag{2.17}$$

This equation splits the difference between the summary measures M(1) and M(2) into two components:

1. $M(1) - M_1(2)$, which can be rewritten as $M_1(1) - M_1(2)$. This component measures that part of the overall difference $M(1) - M(2)$ which is attributable to ***differences in measures specific for compositional categories***, since it measures the difference between the summary measures with differences in composition controlled (directly standardized) for.
2. $M_1(2) - M(2)$. This component measures that part of the overall difference which is left after controlling for differences in composition. In other words it is the component which is attributable to differences in composition.

In similar fashion, if we adopt P(2) as the standard population and calculate a directly standardized value of M(1) (i.e., $M_2(1)$), then it is clear that:

$$M(1) - M(2) = (M(1) - M_2(1)) + (M_2(1) - M(2)) \tag{2.18}$$

This equation likewise splits the difference between M(1) and M(2) into two components:

1. $M_2(1) - M(2)$, which can be rewritten as $M_2(1) - M_2(2)$, also measures that part of the overall difference $M(1) - M(2)$ which is attributable to differences in measures specific for compositional categories. Again it measures the difference between the summary measures with differences in composition controlled (directly standardized) for.
2. $M(1) - M_2(1)$, which again measures that part of the overall difference which is left after controlling for differences in composition. In other words it, too, measures the component of the overall difference which is attributable to differences in composition.

We have obviously derived ***two*** ways of measuring the component of the difference between M(1) and M(2) which is due to differences in measures specific for compositional categories, and ***two*** ways of measuring the component attiributable to differences in composition. Each pair of expressions is likely to give slightly different results, in the same way as choosing a different standard population earlier gave rise to slightly different standardized ratios. Hence the usual practice is to take averages. The component of the difference M(1) − M(2) that is due to differences in composition-specific measures is given by:

$$C_{SM} = 0.5\ (M(1) - M_1(2) + M_2(1) - M(2)) \tag{2.19}$$

and the component due to differences in population composition is given by:

$$C_C = 0.5\ (M(1) - M_2(1) + M_1(2) - M(2)) \tag{2.20}$$

Where M(1) and M(2) are the summary measures for populations 1 and 2; $M_1(2)$ is the summary measure for population 2 directly standardized to the composition of population 1; and $M_2(1)$ is the summary measure for population 1 directly standardized to the composition of population 2.

Let us now resurrect yet again our example involving CDRs for Malaysia and Australia. In the notation we were using with that example:

$$M(1) - M(2) = D(1) - D(2) = 4.84 - 7.16 = -2.32$$

This is the difference between the two CDRs. The values of elements required to evaluate Eqs. 2.19 and 2.20 which we obtained earlier in this chapter were:

$$D(1) = 4.84;\ D(2) = 7.16;\ D_1(2) = 3.19;\ D_2(1) = 10.42$$

From Eq. 2.19 we have that:

$$\begin{aligned} C_{SM} &= 0.5\,(D(1) - D_1(2) + D_2(1) - D(2)) \\ &= 0.5\,(4.84 - 3.19 + 10.42 - 7.16) \\ &= 2.455 \end{aligned}$$

And from Eq. 2.20 we have that:

$$\begin{aligned} C_C &= 0.5\ (D(1) - D_2(1) + D_1(2) - D(2)) \\ &= 0.5\,(4.84 - 10.42 + 3.19 - 7.16) \\ &= -4.775 \end{aligned}$$

The first thing we will note is that:

$$C_{SM} + C_C = 2.455 + (-4.775) = -2.32 = D(1) - D(2)$$

This is what we expect; the sum of the two components C_{SM} and C_C should always equal the difference between the two summary measures we are decomposing. If this is not the case it is a clear sign that something is wrong and that calculations should be checked.

But how do we interpret our results? We examine the components (2.455 and −4.775) in relation to the difference between the two summary measures (−2.32) ***paying particular attention to their signs***. What is important is ***not*** the sign on a component itself (i.e., whether it is positive or negative), but ***how that sign compares with the sign on the overall difference between the two summary measures***. If the sign on a component is ***the same as*** that on the overall difference, that component ***helped produce*** the overall difference. If the sign on a component is ***opposite to*** that on the overall difference, that component has ***partially offset, or moderated*** the overall difference (i.e., made it less substantial than it would otherwise have been).

In our example, the difference between the two summary measures is negative because we happened to designate the population with the smaller CDR (Malaysia) population 1 and that with the larger CDR (Australia) population 2. The negative difference equates with a higher CDR for Australia. Any component with a negative sign helped to produce this outcome. The specific measures component C_{SM} has the ***opposite*** sign to the difference between the summary measures; this means that differences in composition-specific measures operated in the ***opposite*** direction – that is, in the direction of producing a higher CDR not for Australia, but for Malaysia. As this component on its own indicates the direction of the differential in the ***underlying*** levels of mortality in the two countries, we confirm earlier findings that the underlying level of mortality was in fact higher in Malaysia. The other component, C_C, capturing differences in population composition, has a negative sign. What is more, its absolute value (i.e., its numeric value disregarding the negative sign) is as large as those of the C_{SM} component and the difference between the summary measures combined. This means that differences in population composition both ***cancelled out*** the effect of differences in composition-specific measures ***and*** produced the overall differential in the CDR favouring Australia. The two components operated in opposite directions, one favouring a higher CDR for Malaysia and the other a higher CDR for Australia. The latter (composition) component was the stronger, and the overall difference between the summary measures reflects this.

What if we had designated Australia as population 1 and Malaysia as population 2? The difference $D(1) - D(2)$ would have been $7.16 - 4.84 = 2.32$, and the two components would have had the values $C_{SM} = -2.455$ and $C_C = 4.775$. In other words, all the quantities of interest would have had the same ***numeric*** values, but their signs would have been the opposite of what they were above. Our conclusion would, however, be unaffected. Remember, it is not the signs themselves that matter, but ***how they compare***. The signs on components can only be interpreted ***relative to the sign on the overall difference between our summary measures*** (in this case relative to the sign on $D(1) - D(2)$). The sign on C_{SM} is still opposite to that on $D(1) - D(2)$ (they are now just respectively negative and positive, instead of positive and negative), and the sign on C_C is still the same as that on $D(1) - D(2)$ (they are just both now positive, instead of both being negative). We are still going to conclude that the C_{SM} component favoured a higher CDR for Malaysia, but was cancelled out and more by the C_C component working in the opposite direction, giving rise to an overall differential in favour of Australia. Clearly ***it doesn't matter*** which way round we designate the populations as populations 1 and 2. We arrive at the same conclusion.

We noted earlier that each of our components could be measured in ***two*** ways, and that the answers each gave were likely to be a little different. As a consequence, in Eqs. 2.19 and 2.20 we ***averaged*** the two expressions for each component. The phenomenon which produces (generally slight) inequalities between the answers obtained from the alternate ways of measuring each component is known as ***interaction***. It is given, in the general case, by the equation:

$$\mathrm{I} = 0.5\,(\mathrm{M}(1) - \mathrm{M}_1(2) - \mathrm{M}_2(1) + \mathrm{M}(2)) \tag{2.21}$$

One way of viewing the interaction term I is as a measure of the degree of faith we can have in C_{SM} and C_{C}. We are saying that the precise values of C_{SM} and C_{C} are uncertain, but that they respectively lie somewhere in the ranges:

$$\mathrm{C}_{\mathrm{SM}} \pm \mathrm{I} \text{ and } \mathrm{C}_{\mathrm{C}} \pm \mathrm{I}$$

From a practical point of view the important consideration is whether possible variations in the ‘true’ values of C_{SM} and C_{C} within the ranges defined here raise any doubts over the general conclusion we have drawn. We need to bear in mind that not all combinations of ‘true’ values of C_{SM} and C_{C} in these ranges are possible; once the ‘true’ value of one component has been determined the ‘true’ value of the other is fixed by the requirement that $\mathrm{C}_{\mathrm{SM}} + \mathrm{C}_{\mathrm{C}}$ equals the difference between the two summary measures that is being decomposed. We could not, for example, have $\mathrm{C}_{\mathrm{SM}} + \mathrm{I}$ in combination with $\mathrm{C}_{\mathrm{C}} + \mathrm{I}$; whatever is added to or subtracted from C_{SM} to get its ‘true’ value must in turn be subtracted from or added to C_{C}. Usually, however, ‘true’ values do not alter our general conclusion, since I tends to be small in comparison with C_{SM} and C_{C}. It is nevertheless a wise precaution to check your interaction term.

In our Malaysia/Australia example the interaction term has the value:

$$\begin{aligned} \mathrm{I} &= 0.5\ \ (\mathrm{D}(1) - \mathrm{D}_1(2) - \mathrm{D}_2(1) + \mathrm{D}(2)) \\ &= 0.5\,(4.84 - 3.19 - 10.42 + 7.16) \\ &= -0.805 \end{aligned}$$

Hence the ‘true’ values of C_{SM} and C_{C} lie in the ranges:

$$\begin{aligned} &\mathrm{C}_{\mathrm{SM}}:\ \ 1.65 \text{ to } 3.26 \\ &\mathrm{C}_{\mathrm{C}}: -3.97 \text{ to} - 5.58 \end{aligned}$$

While these ranges are not especially narrow, having regard to the requirement noted above that only combinations of values such that $\mathrm{C}_{\mathrm{SM}} + \mathrm{C}_{\mathrm{C}} = -2.32$ are valid, no major alteration to our earlier interpretation of the C_{SM} and C_{C} values we calculated is called for. In this example, in the limiting cases, $\mathrm{C}_{\mathrm{SM}} = 1.65$ would pair with

$C_C = -3.97$ and $C_{SM} = 3.26$ would pair with $C_C = -5.58$, with intermediate values of C_{SM} paring with intermediate values of C_C.

Decomposition by Reverse Subtraction

The decomposition approach outlined above requires two direct standardizations to be carried out before it can be implemented. ***Reverse subtraction*** is an alternative approach to decomposition which does not have this requirement and which also is very useful in carrying out more complex decompositions (i.e., decompositions into three or more components). We will, however, stick with the simple two component situation in outlining the approach. We will also couch the discussion in terms of ***change in the value of a summary measure over time for a single population***, although the approach is equally applicable to decomposing the difference between summary measures for two populations at a point in time.

Suppose a summary measure for a population has the values V(1) at some earlier date and V(2) at some later date. Making use of Eq. 2.3, these measures can be written in the form:

$$V(1) = \Sigma_c\, m_1(c)\ .\ p_1(c) \tag{2.22}$$

$$V(2) = \Sigma_c\, m_2(c)\ .\ p_2(c) \tag{2.23}$$

Where c denotes a population composition category; $m_1(c)$ and $m_2(c)$ are specific measures for compositional categories at times 1 and 2; $p_1(c)$ and $p_2(c)$ are proportions of the populations comprising the denominators of V(1) and V(2) in compositional categories at times 1 and 2.

Change in the measure V between time 1 and time 2 is given by:

$$V(2) - V(1) = \Sigma_c\, m_2(c)\ .\ p_2(c) - \Sigma_c\, m_1(c)\ .\ p_1(c) \tag{2.24}$$

Suppose we define quantities $\Delta m(c)$ and $\Delta p(c)$ ($\Delta =$ 'delta') representing the ***changes*** in m(c) and p(c) values between time 1 and time 2. Thus:

$$\Delta m(c) = m_2(c) - m_1(c) \tag{2.25}$$

$$\Delta p(c) = p_2(c) - p_1(c) \tag{2.26}$$

We can rearrange Eqs. 2.25 and 2.26 to give:

$$m_2(c) = m_1(c) + \Delta m(c) \tag{2.27}$$

$$p_2(c) = p_1(c) + \Delta p(c) \tag{2.28}$$

In other words, our m(c) and p(c) values at time 2 equal the sums of (i) their values at time 1 and (ii) the changes in their values between times 1 and 2.

We can now return to Eq. 2.24 and substitute for $m_2(c)$ and $p_2(c)$ from Eqs. 2.27 and 2.28. We get:

$$\begin{aligned} V(2) - V(1) &= \Sigma_c\,(m_1(c) + \Delta_m(c))\ \ (p_1(c) + \Delta p(c)) - \Sigma_c\, m_1(c).p_1(c) \\ &= \Sigma_c\,(m_1(c).p_1(c) + m_1(c).\Delta p(c) + \Delta m(c).p_1(c) \\ &\quad + \Delta m(c).\Delta p(c)) - \Sigma_c\, m_1(c).p_1(c) \\ &= \Sigma_c\, m_1(c).\Delta p(c) + \Sigma_c\, p_1(c).\Delta m(c) + \Sigma_c\, \Delta m(c).\Delta p(c) \end{aligned} \tag{2.29}$$

We have now decomposed the change in summary measure V into three components:

1. A component $\Sigma_c\, m_1(c).\Delta p(c)$ due to ***changes in population composition***. This is what $\Delta p(c)$ refers to – change in population composition.
2. A component $\Sigma_c\, p_1(c)$. $\Delta m(c)$ due to ***changes in composition-specific measures***. This is what $\Delta m(c)$ refers to – change in specific measures.
3. An ***interaction*** component, identifiable by the fact that it ***incorporates more than one delta (Δ) item***. This is a general rule with component analysis by reverse subtraction – the substantive components are expressions featuring ***single*** delta items; expressions featuring multiple delta items are interaction components, or terms.

To obtain Eq. 2.29 we substituted for $m_2(c)$ and $p_2(c)$ in Eq. 2.24. We could equally have used Eqs. 2.25 and 2.26 to find expressions for $m_1(c)$ and $p_1(c)$ and substituted these instead. You may like to go through this exercise. It should lead you to the following alternative equation for V(2) - V(1):

$$V(2) - V(1) = \Sigma_c\, m_2(c).\Delta p(c) + \Sigma_c\, p_2(c).\Delta m(c) - \Sigma_c\, \Delta m(c).\Delta p(c) \tag{2.30}$$

As in Eq. 2.29, we have decomposed the change in summary measure V into three components associated with changes in population composition and in composition-specific measures, and with interaction. We can, however, proceed to a much simpler expression of the components of change; one which has the added benefit that the interaction terms disappear. Because the righthand sides of Eqs. 2.29 and 2.30 both equal V(2) – V(1), we can say that ***half their sum*** also equals V(2) – V(1). Thus:

$$
\begin{aligned}
\mathrm{V}(2)-\mathrm{V}(1) &= 0.5\Big[\Sigma_\mathrm{c}\,\mathrm{m}_1(\mathrm{c})\,.\Delta\mathrm{p}(\mathrm{c}) + \Sigma_\mathrm{c}\,\mathrm{p}_1(\mathrm{c})\,.\Delta\mathrm{m}(\mathrm{c}) + \Sigma_\mathrm{c}\,\Delta\mathrm{m}(\mathrm{c})\,.\Delta\mathrm{p}(\mathrm{c}) \\
&\quad + \Sigma_\mathrm{c}\,\mathrm{m}_2(\mathrm{c})\,.\Delta\mathrm{p}(\mathrm{c}) + \Sigma_\mathrm{c}\,\mathrm{p}_2(\mathrm{c})\,.\Delta\mathrm{m}(\mathrm{c}) - \Sigma_\mathrm{c}\,\Delta\mathrm{m}(\mathrm{c})\,.\Delta\mathrm{p}(\mathrm{c})\Big] \\
&= 0.5\,\left[\Sigma_\mathrm{c}\,(\mathrm{m}_1(\mathrm{c}) + \mathrm{m}_2(\mathrm{c}))\,.\Delta\mathrm{p}(\mathrm{c}) + \Sigma_\mathrm{c}\,(\mathrm{p}_1(\mathrm{c}) + \mathrm{p}_2(\mathrm{c}))\,.\Delta\mathrm{m}(\mathrm{c})\right] \\
&= \Sigma_\mathrm{c}\,\overline{\mathrm{m}}(\mathrm{c})\,.\Delta\mathrm{p}(\mathrm{c}) + \Sigma_\mathrm{c}\,\overline{\mathrm{p}}(\mathrm{c})\,.\Delta\mathrm{m}(\mathrm{c}) \qquad (2.31)
\end{aligned}
$$

Where $\overline{\mathrm{m}}(\mathrm{c})$ and $\overline{\mathrm{p}}(\mathrm{c})$ are the ***averages*** of the m(c) and p(c) values at times 1 and 2 for each compositional category. In other words:

$$
\begin{aligned}
\overline{\mathrm{m}}(\mathrm{c}) &= 0.5\;(\mathrm{m}_1(\mathrm{c}) + \mathrm{m}_2(\mathrm{c})) \\
\overline{\mathrm{p}}(\mathrm{c}) &= 0.5\;(\mathrm{p}_1(\mathrm{c}) + \mathrm{p}_2(\mathrm{c}))
\end{aligned}
$$

We now have the change in a summary measure V decomposed into two components, one associated with changes in population composition (the Δp(c) component) and the other associated with changes in specific measures (the Δm(c) component). Equation 2.31 is the one you would use in working through a real example, and if doing the calculations by hand it would again make sense to draw up a calculating table. This (or an excel spreadsheet) would have columns headed:

$\mathrm{m}_1(\mathrm{c})$; $\mathrm{m}_2(\mathrm{c})$; $\mathrm{p}_1(\mathrm{c})$; $\mathrm{p}_2(\mathrm{c})$; $\overline{\mathrm{m}}(\mathrm{c})$; $\Delta\mathrm{m}(\mathrm{c})$; $\overline{\mathrm{p}}(\mathrm{c})$; $\Delta\mathrm{p}(\mathrm{c})$; $\overline{\mathrm{m}}(\mathrm{c})\,.\Delta\mathrm{p}(\mathrm{c})$; $\overline{\mathrm{p}}(\mathrm{c})\,.\Delta\mathrm{m}(\mathrm{c})$

Values of the two components would be found by summing the final two columns. These should then be added and checked against the value of V(2) − V(1), which their sum is supposed to equal. This provides a quick check to see whether some simple mechanical error has been made.

When we use this methodology to decompose change over time it is obvious that V(2) should relate to the later point in time. When using it to decompose a difference between two populations at a single point in time the decision as to which population should be population 1 and which population 2 is again arbitrary. The main thing is to be consistent. Once a designation has been made, stick with it. Make sure all the V(1), $\mathrm{m}_1(\mathrm{c})$ and $\mathrm{p}_1(\mathrm{c})$ values relate to one population and all the V(2), $\mathrm{m}_2(\mathrm{c})$ and $\mathrm{p}_2(\mathrm{c})$ values relate to the other one. Switching designations and mixing up values will cause all sorts of strife.

It was mentioned above that the reverse subtraction approach to decomposition can be extended to deal with more complex situations than the two-component one discussed here. The following is a recipe for applying the approach in any situation.

Step 1: If the measure to be decomposed is denoted by V, obtain an expression (equation) for V in terms of the factors you wish to recognize in the decomposition. This expression will invariably involve summing terms over a series of compositional categories of some sort.

Step 2: Write the difference, or change over time, in the measure V (i.e., V(2) – V(1)) in terms of the expression for V obtained at step 1. Elements in the resulting equation should take the form x(1) where they relate to population, or time, 1 and x(2) where they relate to population, or time, 2, where x is any variable recognized on the righthand side of the equation for V.

Step 3: Rewrite each element in the expression for V(2) in the form $x(2) = x(1) + \Delta x$, where Δx is the difference, or change, in x between populations, or times, 1 and 2 (i.e., $\Delta x = x(2) - x(1)$).

Step 4: Expand the resulting expression for V(2) – V(1) and collect together terms which feature the same combinations of Δ (delta) items.

Step 5: Go back to the expression for V(2) – V(1) obtained at step 2 and rewrite each element in the expression for V(1) in the form $x(1) = x(2) - \Delta x$ (note the negative sign).

Step 6: Again expand the resulting expression for V(2) – V(1) and collect together terms which feature the same combinations of Δ items.

Step 7: Add together the expanded expressions for V(2) – V(1) obtained at steps 4 and 6, then divide by 2 and simplify where possible.

In the expression obtained for V(2) – V(1) at step 7, those components which are functions of ***single*** Δ items are the components of the overall difference or change due to that particular factor (i.e., the Δx component is the component due to factor x). Any component which is a function of ***two or more*** Δ items is a separate interaction factor. Such components are commonly added together to form a single (generally small) overall interaction component.

The data presented in Table 2.3 are results from a reverse subtraction decomposition in which the mechanics of a continuous 25-year increase in the 'illegitimacy ratio' for New Zealand's non-Maori population were investigated (Carmichael, 1985). Although the example is dated its value in illustrating how to go about interpreting a set of decomposition results is considerable. The nowadays rather quaintly named 'illegitimacy ratio' measures the percentage of all live births in a population that are non-marital; that is, it is the ratio of non-marital live births to total (marital plus non-marital) live births. It is a measure whose value can change in response to change in four factors: in age-specific rates at which unmarried women bear children; in age-specific rates at which married women bear children; in the propensity of women of reproductive age to be married at different ages; and in the age structure of the female population of reproductive age. To perform the decomposition it was necessary to obtain an equation in terms of these four factors, and then to follow the steps outlined above. The equation used was:

$$R = (\Sigma_{x=15,49}\, f_x.q_x.i_x) \,/\, [\Sigma_{x=15,49} f_x.q_x.i_x + \Sigma_{x=15,49}\, f_x.\,(1 - q_x)\,.j_x] \qquad (2.32)$$

Where R = illegitimacy ratio; x = a single-year age group; f_x = the proportion of all females of childbearing age aged x; q_x = the proportion of females aged x not currently married; i_x = the non-marital fertility rate per unmarried woman for females aged x; j_x = the marital fertility rate per currently married woman for females aged x.

Table 2.3 Decomposition of change in the New Zealand Non-Maori illegitimacy ratio, 1951–1976

	1951–1962	1962–1966	1966–1971	1971–1976
Starting ratio (R(1))	4.38	6.51	9.49	11.23
Finishing ratio (R(2))	6.51	9.49	11.23	13.83
Total change (R(2) − R(1))	2.13	2.98	1.74	2.60
Change due to:				
Non-marital fertility rates	4.19	1.04	1.39	−1.64
Marital fertility rates	−0.15	1.64	0.87	3.42
Marriage pattern	−1.45	−0.28	−0.28	1.31
Age structure	0.34	0.39	−0.29	−0.39
Interaction	−0.81	0.19	0.05	−0.10

Source: Carmichael (1985: 170)

The algebra involved in expanding Eq. 2.32 in accordance with steps 1–7 above was complex and is not reproduced here. To interpret the results obtained (Table 2.3) we need first to ascertain what the signs (+ or −) on the various components mean. For each of the four periods for which results are presented the illegitimacy ratio increased; the ‘total change’ figures, found by subtracting the ‘starting ratio’ from the ‘finishing ratio’ (R(2) − R(1)), are all positive. In line with the principle previously enunciated for interpreting components, positive components, having the same sign as the ‘total change’ figures, contributed to those changes; i.e., they contributed to the rise in the illegitimacy ratio over the relevant period. Negative components partially offset, or moderated, this upward momentum.

Under what circumstances would we expect each component to be causing the illegitimacy ratio to rise, and under what circumstances would we expect each to be causing it to fall?

1. Other things remaining equal, increasing ***non-marital fertility rates*** will boost the number of non-marital compared to marital births, and so tend to increase the illegitimacy ratio. A positive ‘non-marital fertility rates’ component thus attests to rising non-marital fertility rates tending to raise the illegitimacy ratio; a negative component points to falling non-marital fertility rates tending to lower it.
2. Other things remaining equal, increasing ***marital fertility rates*** will boost the number of marital compared to non-marital births and, as marital births contribute only to its denominator, tend to lower the illegitimacy ratio. Thus, a positive ‘marital fertility rates’ component attests to falling marital fertility rates tending to raise the illegitimacy ratio; a negative component points to rising marital fertility rates tending to lower it.
3. A change in the ***marriage pattern*** that results in women of reproductive age being less likely to be married increases exposure to the risk of giving birth while unmarried and thus tends to raise the illegitimacy ratio. A positive ‘marriage

pattern' component is an indication that such a change occurred (ages at marriage rose, the propensity to marry fell, the rate of divorce increased, the rate of remarriage following divorce fell, or a combination of these things happened) and tended to raise the illegitimacy ratio; a negative component indicates an increased tendency for women of reproductive age to be married, a consequent reduced exposure to non-marital (and increased exposure to marital) childbearing, and hence a tendency to lower the illegitimacy ratio.

4. A change in ***age structure*** among women of reproductive age is likely to result in a rise in the number of non-marital compared to marital births if it renders the younger reproductive ages (15–19 years and to some extent 20–24 years) more important. This is so because, regardless of age-specific ***rates*** of non-marital childbearing, these age groups provide the overwhelming bulk of women who are unmarried (mostly ***never*** married) and thus exposed to the risk of such childbearing. A positive 'age structure' component is likely to mean that an increase in the importance of the younger reproductive ages has tended to raise the illegitimacy ratio; a negative component probably indicates that the importance of these age groups declined, and tended to lower the illegitimacy ratio.

With this background we can proceed to interpret Table 2.3. Between 1951 and 1962, rising non-marital fertility rates were overwhelmingly the main reason for the rise in the illegitimacy ratio. Indeed the 'non-marital fertility rates' component was almost twice as large as the 'total change'. It was offset principally by changes in the marriage pattern unfavourable to non-marital childbearing. This period coincided with the post-war marriage 'boom', with ages at marriage falling and the propensity to ever marry increasing. These trends tended to reduce the period of theoretical exposure to the risk of non-marital childbearing, but they could not completely offset rising levels of sexual activity outside marriage and the increased non-marital fertility rates that resulted.

Between 1962 and 1966 the pattern changed. Non-marital fertility rates continued to rise and therefore to boost the illegitimacy ratio, but they were not the dominant force they had been earlier. Instead, falling marital fertility rates associated with the advent of new contraceptive technology (principally the oral contraceptive pill) in the early 1960s assumed this mantle. As during 1951–1962, an increasing importance of the younger reproductive age groups was a minor factor in the illegitimacy ratio rising.

A somewhat similar pattern prevailed between 1966 and 1971, although rising non-marital fertility rates regained the status of dominant force from falling marital fertility rates, and age structure changes became marginally *un*favourable to the upward trend. Then, between 1971 and 1976 a totally new causative pattern emerged. Having for so long been an influence for increase in the illegitimacy ratio, non-marital fertility rates fell and became a force for decline. This was a period when induced abortion became more accessible, and with informal cohabitation also proliferating it is possible that unmarried women became contraceptively more adept as well. However, this reversal was more than compensated for by accelerated marital fertility decline and a sharp turnaround in the role played by the ‘marriage pattern’ component as the marriage ‘boom’ ended and ages at marriage began to rise. Both trends were partly attributable to the aforementioned improved access to abortion, which saw marriages precipitated by bridal pregnancy, and consequently early marital fertility, both fall dramatically.

This example serves to graphically illustrate that considerable complexity may underlie what on the surface appears to be a straightforward, ongoing secular trend. Decomposition techniques can greatly enhance our understanding of such demographic trends.

Reference

Carmichael, G. A. (1985). Non-marital pregnancies in New Zealand since the second world war. *Journal of Biosocial Science, 17*, 167–183.

Chapter 3
The Cohort and Period Approaches to Demographic Analysis

Conceptualizing Problems in Demographic Analysis: Lexis Diagrams

As the title to this chapter suggests, demographers adopt either (or both) of two perspectives, or approaches, in analysing demographic data. They are ***the cohort approach*** to demographic analysis and ***the period approach*** to demographic analysis. The difference between these two approaches is most readily understood by making use of a graphical device known as ***the Lexis diagram***. Named for Wilhelm Lexis (1837–1914), a German statistician, actuary and economist who was the first to use it, the Lexis diagram is also an invaluable tool for conceptualizing the construction of demographic measures and the procedures followed in demographic estimation and projection. The first part of this chapter is devoted to describing it and the concepts on which its use rests.

Time and Its Straight Line Representation

For purposes of analysis, demographers must generally ***classify*** their data according to certain criteria, and often those criteria involve some concept of ***time***. Demographic ***events*** (births, deaths, migratory moves, marriages, divorces, etc.) are frequently classified by ***age*** or some sort of ***duration*** (of marriage, divorce, residence, etc.), and are almost always classified by ***period*** (usually calendar year) ***of occurrence***. ***Populations*** are also frequently classified by ***age*** or some sort of ***duration***, and also by ***date of enumeration*** (the date at which the population count was made). For example, deaths are frequently classified by age of the deceased, births by age of mother and/or (in the case of marital births) duration of marriage, divorces by duration of marriage; all three types of event are also classified by year

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_3

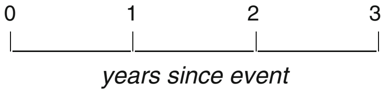
of occurrence. The female population of reproductive age for a country might be classified by age, its currently married component might be classified by duration of marriage, and both populations would further be classified by date of enumeration.

The Lexis diagram is a device for representing graphically demographic data which are classified according to criteria which involve some concept of time. In order to understand and use Lexis diagrams it is essential to appreciate the distinction between ***time instants*** and ***time intervals***. When our time scale is an age or duration scale (and really an age scale is also a duration scale – age is simply duration, or time elapsed, ***since birth***), the distinction we make is between ***exact ages or durations*** (which are time instants in the lives of individuals) and ***ages or durations in completed years***, or last birthday/anniversary (which are attributes that apply to an individual for a period of a year).

When expressed as whole numbers of years, exact ages or durations are fleeting moments – instants – in the lives of individuals which correspond to birthdays or anniversaries. A person is 'exact age 10' at the time of his/her tenth birthday – that is, ***exactly*** 10 years after (s)he was born; one moment (s)he hasn't yet turned ten, the next (s)he is exactly 10 years old, and the next (s)he is, by a fraction of a second, ***more*** than 10 years old ('exact age 10 years, 0.1 of a second', say). An important example of an exact age is exact age 0, which is attained at the instant of birth. For a married couple, 'exact marriage duration 4 years' is their fourth wedding anniversary. They might spend a whole day celebrating this anniversary, but only for an instant can they say they have been married for exactly 4 years; either side of that instant they have been married for less than 4 years and for more than 4 years.

On the other hand, when you are asked 'How old are you?', typically you will give not your current exact age ('I'm aged 25 years, 10 months, 6 days, 15 h, 23 min, 13.4 s') but the age you attained at your last birthday ('I'm 25 years old'). This age is an attribute that remains with you for a full year (from the time you turn 25 until you turn 26 a year later). Similarly a married person, if asked 'How long have you been married?' will typically respond along the lines 'I've been married for 6 years', meaning 'My spouse and I have passed our sixth wedding anniversary, but have yet to reach our seventh'. These 1-year periods during which a particular age last birthday or duration last anniversary applies to a person are examples of age or duration ***intervals***.

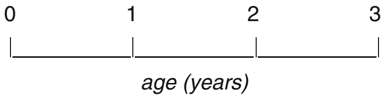
Consider a straight line with a series of points located on it to represent years elapsed since some important event occurred.



The points marked on this line represent time ***instants***; they represent the instant the event occurred and subsequent instants ***exactly*** 1, 2 and 3 years after it occurred. We could also represent other time instants on the line. For example, the instant exactly 9 months after the event would be represented by a point three-quarters of the way between the scaled points 0 and 1; the instant exactly 2 years 6 months after the event would be represented by a point halfway between the scaled points 2 and 3.

While time ***instants*** are represented by ***points*** on our line, time ***intervals*** are represented by ***segments*** of the line. Thus, for example, the segment stretching from the point representing exactly 1 year since the event to that representing exactly 2 years since the event corresponds to a time interval of 1 year, throughout which one complete year (but not yet two) had elapsed since the event occurred.

Another way of thinking of our line is as a scale of ***ages***. Age for an individual is simply years elapsed since the event of ***birth*** occurred.



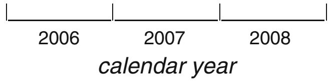
Points marked on this line represent ***exact ages*** or ***birthdays***. We would locate an individual at these points at birth (exact age 0) and when (s)he was exactly 1, 2 and 3 years old. We could equally locate an individual at any other exact age between birth and exact age 3 years by identifying the appropriate point along the line; for example, the point corresponding to exact age 2 years, 5 months, 3 weeks, 2 days, 15 h, 30 min. On the other hand, ***segments*** of the line stretching from one exact age in whole years to the next represent 1-year ***age intervals*** or ***age groups***.

Consider, for instance, the segment stretching from exact age 2 years to, but excluding, exact age 3 years. How might we describe the group of individuals whose exact ages at a point in time were located somewhere along this segment? They are the group of individuals who at that point in time ***had passed their second birthdays, but had yet to reach their third birthdays***. In other words, they are the group of individuals (the population) ***2 years old***, or ***aged two last birthday*** (aged ***two completed years***). Our segment therefore corresponds to (or represents) a ***1-year age group*** within the population.

We could equally think of our line as a scale of ***marriage durations***; that is, a scale of time elapsed since the event of marriage.

Points on the line now represent ***exact marriage durations*** or ***wedding anniversaries***. Segments of the line represent 1-year ***marriage duration groups*** (i.e., groups of married people who were between one wedding anniversary and the next one). We could similarly think of our line as a scale of ***durations of divorce*** (points would represent exact durations of divorce or anniversaries of divorce; segments would represent 1-year duration of divorce groups), as a scale of ***durations of residence*** (points would correspond to exact durations of residence or anniversaries since taking up residence; segments to 1-year duration of residence groups), etc.

Having discussed ***age*** and ***duration*** time scales, we will also look at another time scale – a scale of ***calendar time***.



Again we can think in terms of time ***instants*** and time ***intervals***. On this occasion, because it is more conventional and convenient, we have labelled intervals rather than points on our time line. Those intervals, or segments, represent calendar years. But what do the points represent? They represent time instants defined by ***dates*** and ***hours of the day***. Thus, the four points marked on our line represent midnight on 31 December in each of the years 2005, 2006, 2007 and 2008. We could equally identify other time instants; for example, the date of a census held at midnight on 30 June, 2006 would be represented by a point halfway along the segment which corresponds to calendar year 2006.

The Lexis Diagram

The ideas we have been discussing are the basis of the Lexis diagram**.** ***A Lexis diagram consists of a grid of squares which has its horizontal axis scaled in calendar time and its vertical axis scaled in age or duration time*** (Fig. 3.1). On the horizontal axis time ***intervals*** are usually labelled, normally, although not necessarily, in calendar year units. On the vertical axis you will sometimes find age or duration ***instants*** labelled (i.e., ***exact*** ages or durations), and sometimes age or duration ***intervals*** (i.e., age or duration ***groups***). The vertical axis should have a caption that accurately reflects the way in which it has been scaled (e.g., ‘exact age’, or ‘age group’).

Three types of geometric image are used to represent demographic concepts or quantities on a Lexis diagram: ***points***, ***lines***, and ***areas***.

Points on Lexis Diagrams

A point on the Lexis grid may represent either of two things:

1. ***A demographic event*** classified according to ***date of occurrence*** and the ***exact age or duration*** at which the individual concerned experienced it.
2. ***A member of a population*** classified according to ***date of enumeration*** and ***exact age or duration*** at the time of enumeration.

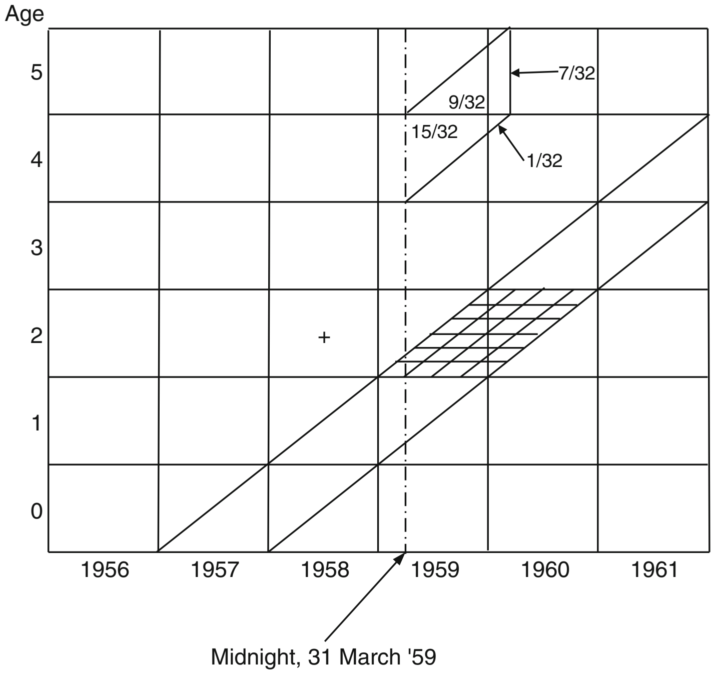


Fig. 3.1 The Lexis diagram

Thus, the point represented by the '+' sign in Fig. 3.1 might represent a ***death*** which occurred at midnight on 30 June, 1958 and involved a person who at the time was exactly 2 years and 6 months old. Alternatively it could represent simply a member of a population enumerated at midnight on 30 June, 1958 who was exactly 2 years and 6 months old at the time of enumeration.

In practical applications of the Lexis diagram we are rarely interested in ***individual*** points. Our interest in points stems from the fact that ***lines*** and ***areas*** on the Lexis grid can be considered to be ***collections*** of points. It is lines and areas that normally interest us.

Lines on Lexis Diagrams

Three types of line on a Lexis diagram are of interest: ***horizontal*** lines, ***vertical*** lines and ***diagonal*** lines running at 45° from bottom left to top right. ***A horizontal line*** represents the group of individuals who attained the exact age or duration defined by the location of the line on the vertical scale (i.e., who reached that stage of the life cycle) during the calendar period defined by the length of the line. Hence, in Fig. 3.1, the line forming the base of the shaded parallelogram represents persons who attained exact age 2 years (i.e., who reached their second birthdays) during calendar year 1959; the line forming the base of the diagonal band through the diagram represents persons who attained exact age 0 (i.e., who were born) during calendar year 1957.

Reference to 'reaching a stage of the life cycle' should ring a bell. We learned in Chap. 1 that the denominator of a demographic probability, in a common method of estimating such measures, is the population at risk at the beginning of the relevant life cycle phase of risk. Thus, while it is possible to conceive of exceptions, ***the denominator of a demographic probability is normally a population quantity we would represent on a Lexis diagram as a horizontal line***.

A vertical line on a Lexis diagram represents the group of individuals who at a time instant (date) defined by the location of the line on the horizontal scale were members of the age or duration group defined by the length of the line. Thus, in Fig. 3.1, the vertical line dividing the shaded parallelogram in half represents persons who at midnight on 31 December, 1959 were aged 2 years (i.e., 2 years last birthday); the portion of the vertical line drawn at midnight on 31 March, 1959 that lies within the diagonal band through the diagram represents persons who at that point in time were aged between exactly 1 year 3 months, and exactly 2 years 3 months.

You may recall, again from Chap. 1, that it is common practice to use the mean, or mid-year, population at risk as the denominator of a demographic rate. The mid-year population will be a population at a particular time instant. ***The denominator of a demographic rate is thus often approximated by a population quantity we would represent on a Lexis diagram as a vertical line***. Placing this statement alongside that highlighted two paragraphs above, we now have a new way of differentiating demographic rates from demographic probabilities – in terms of the ways their denominators are normally, or often, represented on a Lexis diagram.

The third type of line conventionally found on Lexis diagrams, ***a diagonal line*** set at 45° to the horizontal axis, is often referred to as a ***life line***. Life lines trace the passage of individuals through life (Fig. 3.2). A person's life line starts at the date on the horizontal axis on which (s)he attained exact age 0 (i.e., was born). It then runs diagonally upwards to the right at 45° until either the current date or the date the person died is reached. Figure 3.2 shows the life line of a person who was born, let us say, on 31 March 1980, and who thereafter aged 1 year for each year of calendar time that passed until dying around mid-February (let us say on 14 February) 1985, about a month and a half short of his/her fifth birthday (i.e., aged four).

We can also think of life lines in the context of forms of life ***other*** than that which begins at birth and ends at death. For example, we can think in terms of a person's (or a couple's) life ***in a marriage***. That 'life' begins at the date of marriage (at exact marriage duration 0 years – the instant they are pronounced 'man and wife') and ends when the marriage ends (due to either the death of one party or divorce). Similarly we could think of life in the state of ***being a woman of parity two*** (i.e., having had two live births). That 'life' begins at the instant the second live birth occurs (at exact duration in parity two 0 years) and ends when either the woman dies or her third live birth occurs. You might like to reinterpret Fig. 3.2 in terms of

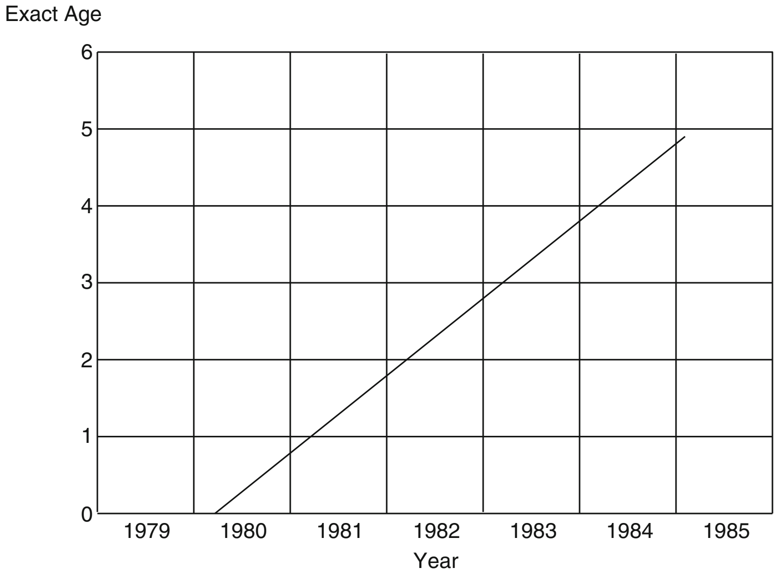


Fig. 3.2 Lexis diagram illustrating concept of a ‘Life Line’

these two forms of ‘life’, with the vertical axis taken first to be a scale of durations of marriage, and then to be a scale of durations in parity two.

From a practical point of view we are rarely concerned with life lines for ***individuals*** when using Lexis diagrams. However, if you look at Fig. 3.3 you can probably appreciate that the ***collection*** of life lines for all people born (or married, or attaining parity two, etc.) in a given calendar year (1980 in this instance) defines a ***diagonal band*** through the Lexis grid. ***All*** life lines for people born in that year lie within this band, and ***only*** life lines for people born in that year lie within it.

You can perhaps also appreciate that any demographic event an individual experiences can be represented by a point lying on his/her life line at the relevant date and exact age. It follows from this and the fact that life lines for individuals born in a given year all lie within a discrete diagonal band that all events occurring to individuals born in that year, when represented as points, lie within the same diagonal band. ***In other words, the diagonal band summarizes, or embraces, the entire demographic experience of persons born during the year in question***. Persons born during a particular year constitute, according to the definition given in Chap. 1, a ***birth cohort*** (a group of individuals who all experienced the demographic event ‘birth’ during a specified period of time). It follows that ***the demographic experience of a birth cohort*** (and indeed of any other type of cohort) ***can be represented as a diagonal band through a Lexis diagram***. Normally we deal with ***calendar-year cohorts***; i.e., cohorts comprising individuals who experienced the demographic event defining the cohort (birth for a birth cohort, marriage for a marriage cohort, etc.) during a calendar year. The diagonal bands representing such

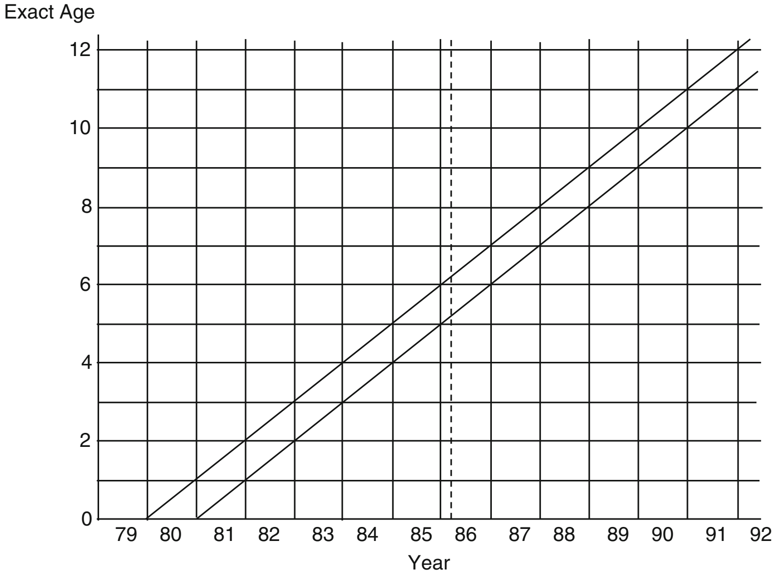


Fig. 3.3 Demographic experience of a cohort represented by a diagonal band through a Lexis diagram

cohorts on a Lexis diagram are of width 1 year, and the diagonal lines defining those bands have common points of intersection with the horizontal and vertical grid lines that are also 1 year apart (Fig. 3.3). However, we do sometimes also deal with ***non-***calendar year cohorts (e.g., cohorts who experienced the defining demographic event between the beginning of July in 1 year and the end of June in the following year), and with cohorts that are wider or narrower than 1 year (e.g., ***5-year*** or ***6-month*** birth cohorts). You might like to try sketching cohorts of these types on a Lexis diagram after the fashion illustrated in Fig. 3.3.

Areas on Lexis Diagrams

Like lines, areas on a Lexis diagram can be thought of as collections of points. Areas on a Lexis diagram represent ***collections of events distributed by exact age/duration at the time of occurrence and date of occurrence***. Put slightly differently, they represent collections of events that occurred ***either*** within a particular age or duration interval over a specified period of time, ***or*** to members of a particular cohort between specified exact ages/durations, specified dates, or a combination of both.

Consider Fig. 3.1 again, and imagine that we have drawn this diagram in connection with an analysis of mortality patterns (i.e., the event to which the diagram pertains is death). The square in which the '+' sign is located represents the collection of deaths at age 2 that occurred during 1958. The shaded parallelogram represents deaths at age 2 of members of the 1957 birth cohort. The triangle defined by the left side of this parallelogram and the '31 March, 1959' vertical line represents deaths at age 2 of members of the 1957 birth cohort which occurred before midnight on 31 March, 1959. Finally, the unshaded parallelogram represents deaths that occurred between 1 April 1959 and 31 March 1960 of members of the birth cohort born between 1 April 1954 and 31 March 1955 (we determine these latter dates either by extending the diagonal lines defining the cohort back to the horizontal axis, or by recognizing that a cohort aged between exact age 4 and exact age 5 at midnight on 31 March 1959 was exact age 0, and therefore born, between 5 and 4 years previously).

Understanding what areas on a Lexis diagram represent can often be aided by defining the area of interest as the intersection of vertical and/or horizontal and/or diagonal bands through the Lexis grid. You can then read off the calendar period and/or age/duration group and/or cohort corresponding to these bands, and this will give you the calendar time-age/duration-cohort dimensions defining the set of events the area represents. For example, in Fig. 3.1 the square in which the '+' sign is located can be thought of as the intersection of the horizontal band corresponding to age group 2 years and the vertical band corresponding to calendar year 1958; hence it represents deaths at age 2 that occurred during 1958. The shaded parallelogram can be thought of as the intersection of the horizontal band corresponding to age group 2 years and the diagonal band representing the 1957 birth cohort; hence it represents deaths at age 2 of members of the 1957 birth cohort. The triangle defined by the left side of this parallelogram and the '31 March, 1959' vertical line can be thought of as the intersection of the same horizontal and diagonal bands and the vertical band corresponding to the period 1 January, 1959 to 31 March, 1959; hence it represents deaths at age 2 of members of the 1957 birth cohort which occurred between 1 January, 1959 and 31 March, 1959 inclusive.

This approach can also be followed ***in reverse*** if the task is to represent a certain set of events on a Lexis diagram. Identify the vertical and/or horizontal and/or diagonal bands that correspond to the calendar time-age/duration-cohort dimensions defining the set of events of interest. The relevant area on the Lexis diagram will then be ***the intersection of these bands***, or if you like, ***the area that is common to all of them***.

Note in passing that a similar approach can also be adopted when trying to understand what ***lines*** on a Lexis diagram represent. Consider, for example, in Fig. 3.1 the portion of the vertical line drawn through 31 March, 1959 that lies within the diagonal band. This line is the intersection of the vertical line representing the population distributed by age at midnight on 31 March 1959 and the diagonal band representing the 1957 birth cohort. Hence it represents the 1957 birth cohort as at midnight on 31 March, 1959 (at which date, by reading off against the vertical axis, we can ascertain that its members were aged between exactly 1 year 3 months and 2 years 3 months). What about the line that forms the base of the shaded parallelogram in Fig. 3.1? It can be thought of as the intersection of the horizontal line corresponding to exact age 2 years with ***either*** the vertical band representing calendar year 1959 ***or*** the diagonal band representing the 1957 birth cohort. Thus it represents ***both*** the population that attained exact age 2 during 1959 ***and*** the 1957 birth cohort at exact age 2, and because we are talking about one and the same line, these two populations are identical.

Returning to areas, though, you might reasonably ask 'Do we always have data on numbers of events neatly classified so as to correspond to areas on the Lexis grid in which we are interested? The answer is emphatically 'No'. ***Invariably our event data are classified so as to correspond to squares on the Lexis grid.*** That is, they are classified according to ***calendar year*** of occurrence and ***age or duration (last birthday or last anniversary)*** at the time of occurrence. How, then, do we estimate numbers of events corresponding to areas on the Lexis grid ***other*** than the grid squares?

A basic assumption underlying the use of Lexis diagrams in demographic analysis is that events occurring, and individuals located, within any time, age or duration interval are distributed EVENLY through that interval. When we translate this assumption onto the two-dimensional Lexis surface we are assuming that ***points representing demographic events which lie in any square of the Lexis grid have an even spread (or density) over that square***.

This enables us to estimate the number of events corresponding to any area on the Lexis surface as follows:

1. Divide the area into segments lying in different squares of the Lexis grid. (If the area lies entirely within ***one*** square this obviously becomes unnecessary.)
2. Determine what fraction, or proportion, of the total area of its grid square each segment occupies.
3. Calculate that fraction, or proportion, of the total number of events represented by the relevant grid square. This gives the estimated number of events represented by the segment.
4. Add the answers obtained for the various segments identified at item 1.

The basis of this procedure is, of course, that if points representing demographic events are evenly distributed over a Lexis grid square, the proportion of those events a segment of the square represents will be identical to the proportion of the square's area occupied by that segment.

By way of illustration we can once again turn to Fig. 3.1. If we again think of this diagram as having been drawn to aid an analysis of mortality***, how many deaths are represented by the unshaded parallelogram?*** Our data almost certainly will consist of deaths classified by single years of age and calendar year of occurrence; that is, they will relate to squares of the Lexis grid. Parts of our parallelogram lie in four different grid squares; hence, in accordance with step 1 of the method just outlined, we divide it into four segments.

1. The lower right segment is a right-angled triangle in which the sides forming the right angle are each one-quarter of the length of the relevant side of the grid square in which the segment lies. Hence, since the area of a triangle is given by area = ½ (base × height), this triangle occupies:

$$\tfrac{1}{4} \times \tfrac{1}{4} \times \tfrac{1}{2} = \tfrac{1}{32} \text{ of the area of the grid square}$$

2. The upper left segment is also a right-angled triangle, but with the sides forming the right angle each three-quarters of the length of the relevant side of the grid square in which the segment lies. Hence the triangle occupies:

$$\tfrac{3}{4} \times \tfrac{3}{4} \times \tfrac{1}{2} = \tfrac{9}{32} \text{ of the area of the grid square}$$

3. The lower left segment can be thought of in either of two ways: as the sum of a rectangle of dimension $^1/_4$ by $^3/_4$ of a grid square side and a right-angled triangle the same size as the one in item 2 above; or as a rectangle of dimension $1 \times ^3/_4$ of a grid square side minus a right-angled triangle the same size as the one in item 2 above. Hence this segment occupies:

$$\left(\tfrac{1}{4} \times \tfrac{3}{4}\right) + \left(\tfrac{3}{4} \times \tfrac{3}{4} \times \tfrac{1}{2}\right) = \tfrac{3}{16} + \tfrac{9}{32}$$
$$= \tfrac{15}{32} \text{ of the area of the grid square}$$

or:

$$\left(1 \times \tfrac{3}{4}\right) - \left(\tfrac{3}{4} \times \tfrac{3}{4} \times \tfrac{1}{2}\right) = \tfrac{3}{4} - \tfrac{9}{32}$$
$$= \tfrac{15}{32} \text{ of the area of the grid square}$$

4. The upper right segment can likewise be thought of as the sum of a rectangle and a right-angled triangle or the difference between a larger rectangle and a right-angled triangle. Hence this segment occupies:

$$\left(\tfrac{1}{4} \times \tfrac{3}{4}\right) + \left(\tfrac{1}{4} \times \tfrac{1}{4} \times \tfrac{1}{2}\right) = \tfrac{3}{16} + \tfrac{1}{32}$$
$$= \tfrac{7}{32} \text{ of the area of the grid square}$$

or:

$$\left(1 \times \tfrac{1}{4}\right) - \left(\tfrac{1}{4} \times \tfrac{1}{4} \times \tfrac{1}{2}\right) = \tfrac{1}{4} - \tfrac{1}{32}$$
$$= \tfrac{7}{32} \text{ of the area of the grid square}$$

Therefore, if deaths at age x during year y are denoted by $D_{x,y}$, the estimated number of deaths represented by the unshaded parallelogram is given by:

$$\text{Deaths} = \tfrac{15}{32} D_{4,59} + \tfrac{9}{32} D_{5,59} + \tfrac{1}{32} D_{4,60} + \tfrac{7}{32} D_{5,60}$$

Having demonstrated how Lexis diagrams can be used under the assumption that events occurring within any time or age/duration interval are evenly distributed through that interval, it is appropriate to warn that this ***is*** an assumption, and there are occasions when it is ***patently invalid***. The classic example concerns infant deaths, which typically do not even remotely approach being evenly distributed by exact age between exact ages 0 years and 1 year. Instead they are very heavily concentrated towards the lower end of this age range. In such a circumstance, where the key underlying assumption is clearly seriously violated, it is not acceptable to go blindly on using a Lexis diagram as a tool of demographic analysis. That said, most of the time the assumption of events being evenly distributed through time or age/duration intervals is an acceptably reasonable one, and Lexis diagrams can be very useful aids to conceptualizing demographic indices and demographic analysis problems.

The Lexis Diagram in Operation

The primary use of Lexis diagrams is to help demographers visualize how to go about solving problems in demographic analysis. A secondary application is as an illustrative device. We will now examine some examples of these applications.

Example 1

Suppose we wanted to estimate the size of the 1980 birth cohort for a population at the date of a census conducted at midnight on 31 March 1986. To do this we need to know how old the 1980 birth cohort was at the date of the census. In tackling this problem we initially apply an important principle in solving problems using Lexis diagrams. We draw ourselves a Lexis diagram and ***represent on it the information we have been given***. We are interested in the 1980 birth cohort, so we draw on our diagram the diagonal band representing the demographic experience of that cohort. We are also interested in the population enumerated at a census held at midnight on 31 March 1986, so we draw a vertical line at that date to represent the census population. The result is Fig. 3.3, which we used earlier to illustrate the representation of cohorts on Lexis diagrams.

We want to know how old the cohort was at the census. We are thus interested in the ***intersection*** of the cohort band with the census line. Reading off the vertical coordinates of the points of intersection of the life lines defining the extremities of the cohort band with the census line we find that, at the date of the census, persons born in 1980 were aged between exactly 5 years 3 months and 6 years 3 months. How many individuals were in this age group? Our census data will not tell us directly; they will tell us how many people were aged 5 last birthday (i.e., between exact age 5 and just short of exact age 6) and how many were aged 6 last birthday (between exact age 6 and just short of exact age 7). The group we are interested in lies ***partly*** in ***each*** of these age groups. We therefore ask ourselves 'What proportion of 5 year-olds and what proportion of 6 year-olds at the 1986 census were members of the 1980 birth cohort?'

Looking at that segment of the vertical census line that lies within the cohort band we find that it consists of two parts: a lower part that extends three-quarters of the way across the horizontal band representing age group 5 years last birthday, and an upper part that extends one-quarter of the way across the horizontal band representing age group 6 years last birthday. Thus, using the assumption that events or individuals located within a time or age/duration interval are evenly distributed through that interval (in this case by exact age), the census population that belonged to the 1980 birth cohort consists of three-quarters of all 5 year-olds and one-quarter of all 6 year-olds. Given the number of 5 year-olds and the number of 6 year-olds at the census, we could then quite easily work out the size of the 1980 birth cohort at the census.

We will pause here to briefly highlight an idea mentioned in passing above and an important implication of that idea. Lexis diagrams are a way of ***visualizing*** demographic problems, in the most literal sense – a way of 'seeing' them with one's eyes. It is therefore ***absolutely vital*** that they be ***drawn accurately***. You are not going to be able to 'see' that this population of interest is three-quarters of that population for which you have data, or whatever the case may be, unless your Lexis diagram is drawn carefully and accurately. We have referred so far to Lexis grid ***squares***, and to life lines running at 45° to the horizontal axis. The ideal is probably to keep to these prescriptions, but it is also possible (and not wrong) to use a Lexis grid that consists of ***rectangles*** of equal size. This occurs when the intervals between vertical and horizontal grid lines differ. The important thing in this instance is to ensure that the interval between vertical grid lines is ***constant***, and that the interval between horizontal grid lines ***is also constant*** (albeit different). If a Lexis grid consists of rectangles rather than squares, the angle of life lines to the horizontal axis will also vary from 45°. However, all life lines should be ***carefully drawn, straight and parallel***, and should increase ***exactly*** one unit of time on the vertical axis for every one unit of time on the horizontal axis.

Example 2

Examine Fig. 3.4, which deals with infant deaths which occurred to the 2009 and 2010 birth cohorts of a hypothetical population. One thing should immediately strike you about this diagram. For each cohort, numbers of infant deaths during the year of birth and the year following the year of birth are shown, and these numbers are

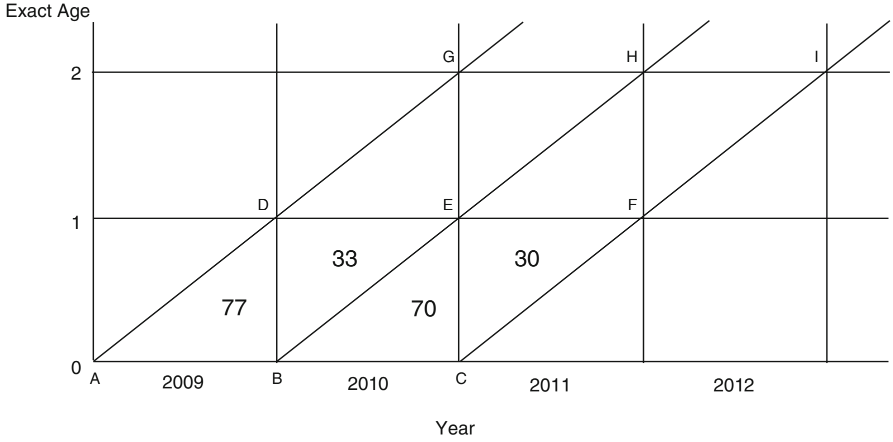


Fig. 3.4 Lexis diagram showing infant deaths to the 2009 and 2010 birth cohorts of a hypothetical population

unequal (a person is at risk of infant death throughout the first year of his/her life, which invariably lies partly in each of two consecutive calendar years). For each cohort the first year of life is represented by a parallelogram (ABED and BCFE) that is subdivided (by lines BD and CE respectively) into two equal-sized triangles. Lexis diagram principles say that the two triangles BED and BCE, which are also of equal size and lie in the same grid square, represent equal numbers of events, but here they do not.

What we have here is a simple representation of the situation that always arises with infant deaths, because they are ***not*** evenly distributed by exact age between exact age 0 and exact age 1. Rather, deaths are heavily concentrated at the lower end of this age range, and as the early part of the first year of life is lived during the year of birth for most individuals, more infant deaths for any birth cohort occur during the year of birth than occur during the following year.

This does not fatally undermine all we have said about the principles that apply in using Lexis diagrams and the usefulness of such diagrams. We could not use a Lexis diagram to split the 103 infant deaths that Fig. 3.4 indicates occurred in 2010 into those involving members of the 2009 birth cohort and those involving members of the 2010 birth cohort. But ***we don't have to***; in this instance we have been ***given*** those data, and the Lexis diagram is serving an illustrative rather than an analytical purpose so far as the distribution of infant deaths among cohort-calendar year segments is concerned.

The line BC in Fig. 3.4 represents births during 2010. Let us say there were 1,000 of them. What do the lines EC and EF represent? EC is the intersection of a vertical line representing the population at midnight on 31 December 2010 and the diagonal band representing the 2010 birth cohort; hence it represents the size of the 2010 birth cohort at the end of 2010, or the number of children born alive during 2010 who remained alive at the end of that year. How do we obtain this number? The size of the cohort at EC is simply its size at BC (which we know equals 1,000) adjusted for the events (deaths) represented by ***the area separating these two lines*** (the triangle BCE, which Fig. 3.4 tells us represents 70 infant deaths); that is, the size of the cohort at EC = 1,000 − 70 = 930 survivors at midnight on 31 December 2010. Note the general principle involved here. ***If we want to know the size of a cohort at a particular date or exact age/duration A, we take its KNOWN size at some other date or exact age/duration B, and ADJUST for those demographic events which altered the cohort's size between A and B. These events are represented by THE AREA SEPARATING the horizontal/vertical lines identifying points A and B in the cohort's life course.***

We can apply this principle again to obtain the size of the 2010 birth cohort at EF from its now known size at EC. EF is the intersection of a horizontal line representing exact age 1 and the diagonal band representing the 2010 birth cohort; hence it represents the size of the 2010 birth cohort at exact age 1. This population is simply the population at EC adjusted for deaths represented by the area separating EC from EF (i.e., by the triangle ECF); that is, the size of the cohort at EF = 930–30 = 900 survivors at exact age 1.

We will take the opportunity here to resurrect an earlier issue, a fuller understanding of which can be obtained with the aid of Fig. 3.4. Recall that in Chap. 1 we noted that the infant mortality rate (IMR) was not a rate but an ***estimate*** of the ***probability*** that a person born during the year for which it was calculated would die in infancy (i.e., before reaching his/her first birthday). We will now use Fig. 3.4 to demonstrate ***why the IMR is a probability***, and ***why it is only an estimate of that probability***. Suppose we denote the population at BC (the 2010 birth cohort at exact age 0) by N_0 (whence the population at EF would be denoted by N_1, that at HI by N_2, etc.). We will also denote the deaths at age 0 of members of the 2010 birth cohort (i.e., the deaths represented by the parallelogram BCFE) by ${}_1D_0$. This introduces a new form of notation that will become very familiar as we proceed to examine life tables. ***Any expression of the form $_nV_x$ means the value of the central variable V for the age/duration group which starts at exact age/duration x and ends at*** (or, in strict terms, marginally short of) ***exact age/duration $x + n$.*** Thus, ${}_1D_0$ denotes deaths (of members of the 2010 birth cohort) between exact age 0 and exact age $0 + 1$ (i.e., exact age 1). We will use it to define a quantity:

$$\mathbf{{}_1q_0 = {}_1D_0/N_0}$$

This quantity is the ratio (for the 2010 birth cohort) of deaths between exact ages 0 and 1 (i.e., deaths at age 0) to the population which attained exact age 0. What sort of measure is it? It is a ***probability*** – the ratio of events (deaths) occurring during a defined phase of the life cycle (the first year following birth) to the population at risk at the beginning of that life cycle phase (the population attaining exact age 0, or the number of live births). ***What*** probability do we have? We have the probability of dying at age 0, which should immediately remind you of the IMR. We said in Chap. 1 that the IMR gave an ***estimate*** of the probability of dying in infancy (i.e., at age 0). In other words, it is an estimate of ${}_1q_0$. So how does the IMR compare with ${}_1q_0$ in Lexis diagram terms?

The 'true' probability of dying at age 0 for the 2010 birth cohort, ${}_1q_0$, is the ratio of deaths in the parallelogram BCFE to the population at BC. The IMR for 2010 is the ratio of deaths at age 0 ***in 2010*** (i.e., deaths in the ***square*** BCED) to the population at BC. Both measures have the same denominator, a measure of the population at risk at the beginning of the life cycle phase of risk (and a population represented on the Lexis diagram by a horizontal line). Both are therefore probabilities. The IMR is an ***estimate*** of the 'true' probability of dying in infancy (for the 2010 birth cohort) because ***it assumes that the infant deaths represented by the triangle BED on the Lexis diagram*** (which actually involved children born the ***previous*** year) ***are a reasonable estimate of the infant deaths represented by the triangle CFE*** (which although occurring in 2011, involved members of the 2010 birth cohort).

Practically speaking, most of the time we have to make do with the IMR as an estimate of ${}_1q_0$. The reason is that we rarely have sufficiently detailed data on infant deaths to enable us to do what is done in Fig. 3.4 – divide those during year y into deaths that involved children born in year y and deaths that involved children born in year y−1. We cannot use a Lexis diagram to help us make this split, because the assumption we would invoke, that deaths at age 0 were evenly distributed between exact ages 0 and 1, is clearly invalid and would produce a completely inaccurate split.

Example 3

We will now examine a more complicated example of the use of a Lexis diagram to aid the calculation of a demographic measure. If you can follow this example, understanding the steps as you go, you probably have a good grasp of the principles underlying the use of Lexis diagrams. Suppose we wished to calculate the following two measures for a population:

1. The probability that a never married woman turning 20 in 1998 married at age 20.
2. The first marriage rate at age 20 for the cohort of never married women turning 20 in 1998.

Suppose further that we have the following data which will prove relevant to our exercise:

Number of never married women aged 19 at 30 June 1998	104,923
Number of never married women aged 20 at 30 June 1998	86,860
Marriages of never married women aged 19 in 1998	10,924
Marriages of never married women aged 20 in 1998	11,848
Marriages of never married women aged 20 in 1999	11,874
Arrivals of never married women aged 19 in 1998	14,701
Arrivals of never married women aged 20 in 1998	15,249
Arrivals of never married women aged 20 in 1999	16,976
Departures of never married women aged 19 in 1998	13,413
Departures of never married women aged 20 in 1998	15,090
Departures of never married women aged 20 in 1999	16,423
Deaths of never married women aged 19 in 1998	74
Deaths of never married women aged 20 in 1998	52
Deaths of never married women aged 20 in 1999	54

To aid us we draw the Lexis diagram shown as Fig. 3.5. There are two general features of this Lexis diagram worthy of note at the outset. First, ***we have drawn only that portion of the Lexis grid that is relevant to our problem***. We have not scaled ages right back to exact age 0 because we are dealing with a cohort's experience at age 20, and most younger ages are irrelevant. There is no point drawing Lexis diagrams that feature masses of grid squares that are never used. We have, in this instance, scaled age from exact age 19 rather than exact age 20; the necessity to do this might not be immediately apparent, but is the sort of thing you should develop a 'feel' for as you become experienced in using Lexis diagrams. Second, ***we have represented on the diagram the information we have been given***. This is invariably a sensible first step in tackling Lexis diagram problems. We want to calculate two measures for the cohort of single women who turned 20 (reached exact age 20) in 1998, so we have identified that cohort on our Lexis diagram. It is the population represented by the horizontal line EB, and we have then drawn the cohort diagonals through E and B. We are also given some data on numbers of never married women by age at 30 June 1998, so we have drawn the vertical dashed line to represent the population of never married women at that date. Finally, we have data on various types of events occurring to never married women aged 19 in 1998, aged 20 in 1998, and aged 20 in 1999; these events are represented by the grid squares ALBE, EBDK and BMCD, which are defined by the intersections of the relevant age groups and calendar time bands through the diagram.

As just noted, the cohort of never married women turning 20 in 1998 is represented on the Lexis diagram by the line EB, and its demographic experience is contained within the diagonal band formed by drawing life lines through this line's end points. According to the definition of a probability previously given, our

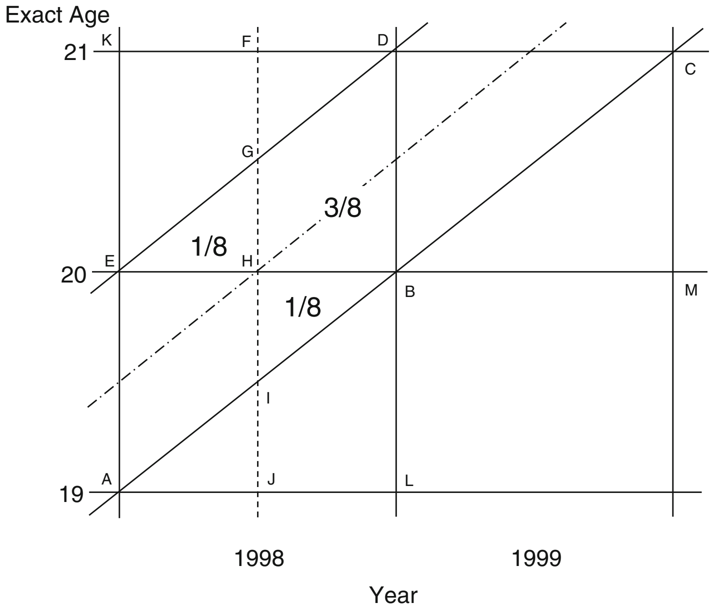


Fig. 3.5 Lexis diagram to aid solution of 'Example 3'

required probability is the ratio of marriages of never married women at age 20 for this cohort to the size of the cohort at the beginning of the life cycle phase of risk:

$$\text{probability} = \frac{\text{marriages of never married women in EBCD}}{\text{never married female population at EB}}$$

Our required rate is the ratio of first marriages at age 20 for the same cohort to the mean size of the cohort during the life cycle phase of risk, our best estimate of which is the never married female population at BD, or at midnight, December 31st, 1998. Hence:

$$\text{rate} = \frac{\text{marriages of never married women in EBCD}}{\text{never married female population at BD}}$$

We thus have three things to estimate: the number of marriages of never married females in EBCD (which provides a common numerator for both measures); the never married female population at EB; and the never married female population at BD. In case you are a little mystified over our specifying the never married female population at BD as the mean population exposed to risk, this mean will be the average of the never married female populations at EB and DC (the beginning and the end of the life cycle phase of risk). Under an assumption that points representing events of a given type which modify the size of the cohort as it ages are evenly distributed over the parallelogram EBCD, this mean will exactly equal the size of the cohort at BD, because numbers of modifying events represented by the triangles EBD and BCD will be exactly equal. Specifying the population at BD as the denominator of our rate also makes intuitive sense in that (i) it is located in the ***middle*** of the 2-year period during which ***some*** members of the cohort of interest are in the relevant age group (recall that the denominator of a rate is often approximated by the ***mid-period*** population at risk), and (ii) it is located at the only point in time at which the population aged 20 is comprised ***entirely*** of members of the cohort of interest. Any earlier and some cohort members were still only 19 while some 20 year-olds had turned 20 in 1997 rather than 1998; any later and some cohort members had by now turned 21 while some 20 year-olds had turned 20 in 1999 rather than 1998.

To estimate the number of marriages of never married women in EBCD: We must make use of the data we have on marriages of never married women, which correspond to squares on the Lexis grid. To use those data we must express parallelogram EBCD in terms of segments, each of which lies in a discrete square. Clearly EBCD is the sum of the two triangles EBD and BCD, each of which comprises half the area of the grid square in which it lies. The two squares, EBDK and BMCD, correspond to events (marriages of never married women) at age 20 in 1998 ($M_{20,98}$) and at age 20 in 1999 ($M_{20,99}$). Thus our required common numerator is:

$$\begin{aligned} {}^{1}\!/_{2}\,(M_{20,98} + M_{20,99}) &= {}^{1}\!/_{2}\ (11{,}848 + 11{,}874) \\ &= 11{,}861 \text{ marriages of never married women} \end{aligned}$$

To estimate the never married female population at EB: We must invoke an important principle in the use of Lexis diagrams. ***We must work from the information we have to the information we want.*** The only data we have on numbers of never married women are the numbers aged 19 and 20 at 30 June 1998. This information is represented on the Lexis diagram by the lines JH and HF respectively. Somehow we must work from this information to obtain an estimate of the number of never married women at EB.

The data we have on numbers of never married women pertain to the date 30 June 1998. Our strategy is to estimate the size of the cohort of never married women of interest at that date (represented by the line GI), then adjust it for events which modify the size of the never married female population over time to obtain the size of the cohort at EB. The important point to appreciate here is that GI and EB ***both represent the size of the never married female cohort we are interested in***; GI just represents it at a particular date, while EB represents it at a particular exact age, or birthday.

We noted earlier that if we want to find the size of a cohort at a particular date or exact age/duration A, we take its known size at some other date or exact age/duration B, and adjust for demographic events which altered the cohort's size between A and

B, these events being represented by ***the area separating*** the horizontal/vertical lines identifying points A and B in the cohort's life course. In this example there is a complication, in that the lines identifying the points in the cohort's life course for which we ***have*** information (GI) and ***want*** information (EB) intersect one another. The easiest way to get around this complication is to divide our cohort into two sub-cohorts by drawing a third diagonal line through the point of intersection of GI and EB (i.e., through H). This effectively divides the cohort of never married women who turned 20 during 1998 into the cohort who turned 20 between January and June 1998 (represented in Fig. 3.5 by the line EH) and the cohort who turned 20 between July and December 1998 (represented by the line HB).

You can perhaps now appreciate that the populations for which we have information (GI) and want information (EB) can each be broken into two segments (GI = GH + HI; EB = EH + HB). Moreover, we can estimate each of the segments of EB from one of the segments of GI. The population at EH is the population at GH adjusted ***backwards*** for relevant demographic events represented by the triangle EHG (the area on the Lexis diagram that separates EH from GH). Similarly the population at HB is the population at HI adjusted ***forwards*** for relevant events represented by the triangle HIB (the area that separates HB from HI). In case these references to adjusting 'backwards' and 'forwards' are a source of confusion, remember that, with the passage of time, individual members of a cohort progress ***up the cohort diagonal band from bottom left to top right***. Thus GH represents the first of our two sub-cohorts at a ***later*** point in the life cycle than does EH, and HI represents the second sub-cohort at an ***earlier*** point in the life cycle than does HB. It follows that we adjust backwards from GH and forwards from HI to obtain EH and HB respectively.

1. The population of never married women represented by EH is the population of never married women represented by GH adjusted for marriages, deaths, in-migrations and out-migrations of never married women represented by the triangle EHG. Since the line GH is half of the line FH, it represents half of the never married female population aged 20 at 30 June 1998:

$$\text{population at GH} = {}^{1}\!/_{2}\, P_{20,98} = {}^{1}\!/_{2}\,(86{,}860) = 43{,}430$$

   As a cohort of never married women ages, it is ***reduced*** in size by marriages, deaths and out-migrations of never married women, and ***increased*** in size by in-migrations of never married women. On this occasion, however, we need to adjust ***backwards*** from the population at GH to obtain the population at EH, so we need to ***add back in*** the marriages, deaths and out-migrations of never married women, and ***subtract*** the in-migrations of never married women represented by the triangle EHG. The triangle EHG lies entirely within the grid square EBDK (representing events that occurred at age 20 during 1998) and occupies one-eighth of the area of that square. Hence the adjustment factor will be:

$$\tfrac{1}{8}\,(M_{20,98}+D_{20,98}+O_{20,98}-I_{20,98}) = \tfrac{1}{8}\,(11{,}848+52+15{,}090-15{,}249)$$
$$= 1{,}467.6 \text{ never married women}$$

and:

$$\text{population at EH} = \text{population at GH} + \text{EHG adjustment factor}$$
$$= 43{,}430 + 1{,}467.6 = 44{,}897.6 \text{ never married women}$$

2. The population of never married women represented by HB is the population of never married women represented by HI adjusted for marriages, deaths, in-migrations and out-migrations of never married women represented by the triangle HIB. Since the line HI is half of the line HJ, it represents half of the never married female population aged 19 at 30 June 1998:

$$\text{population at HI} = \tfrac{1}{2}\,P_{19,98} = \tfrac{1}{2}\,(104{,}923) = 52{,}461.5$$

Again, as a cohort of never married women ages, it is ***reduced*** in size by marriages, deaths and out-migrations of never married women, and ***increased*** in size by in-migrations of never married women. We need to adjust ***forwards*** from the population at HI to obtain the population at HB, so we need to ***subtract*** the marriages, deaths and out-migrations of never married women, and ***add in*** the in-migrations of never married women which are represented by the triangle HIB. The triangle HIB lies entirely within the grid square ALBE (representing events that occurred at age 19 during 1998) and occupies one-eighth of the area of that square. Hence the adjustment factor will be:

$$\tfrac{1}{8}\,(-M_{19,98}-D_{19,98}-O_{19,98}+I_{19,98}) = \tfrac{1}{8}\,(-10{,}924-74-13{,}413+14{,}701)$$
$$= -1{,}213.7 \text{ never married women}$$

and:

$$\text{population at HB} = \text{population at HI} + \text{HIB adjustment factor}$$
$$= 52{,}461.5 - 1{,}213.7 = 51{,}247.8 \text{ never married women}$$

It follows that the required never married female population at EB is:

$$\text{population at EH} + \text{population at HB} = 44{,}897.6 + 51{,}247.8$$
$$= 96{,}145.4 \text{ never married women}$$

***To estimate the never married female population at BD*:** We again must work from known to desired information. We have numbers of never married females corresponding to the lines JH and HF. We can use these to estimate the size of our cohort of never married females at 30 June 1998 (i.e., the population at GI), then adjust for marriage, death, in-migration and out-migration events involving never married women represented by the parallelogram IBDG to obtain the never married female population at BD. This parallelogram is relevant as the area separating the lines GI and BD.

We have that:

$$\begin{aligned}\text{population at GI} &= \text{population at HI} + \text{population at GH}\\ &= \text{half the population at JH} + \text{half the population at HF}\\ &= {}^1\!/_2\,(\mathrm{P}_{19,98} + \mathrm{P}_{20,98})\\ &= {}^1\!/_2\,(104{,}923 + 86{,}860)\\ &= 95{,}891.5 \text{ never married females}\end{aligned}$$

The parallelogram IBDG consists of two segments, the area HBDG which occupies three-eighths of the square EBDK, and the triangle HIB which occupies one-eighth of the square ALBE. In adjusting from the population at GI to that at BD we are adjusting ***forwards*** (i.e., from an earlier to a later point in the cohort's life cycle); hence we need to ***subtract*** marriages, deaths and out-migrations of never married women represented by the parallelogram IBDG and to ***add in*** the in-migrations of never married women it represents. Our adjustment factor is:

$$\begin{aligned}&{}^3\!/_8\,(-\mathrm{M}_{20,98} - \mathrm{D}_{20,98} - \mathrm{O}_{20,98} + \mathrm{I}_{20,98})\\ &\quad + {}^1\!/_8\,(-\mathrm{M}_{19,98} - \mathrm{D}_{19,98} - \mathrm{O}_{19,98} + \mathrm{I}_{19,98})\\ &= {}^3\!/_8\,(-11{,}848 - 52 - 15{,}090 + 15{,}249)\\ &\quad + {}^1\!/_8\,(-10{,}924 - 74 - 13{,}413 + 14{,}701)\\ &= -4{,}403 - 1{,}214\\ &= -5{,}617 \text{ never married women}\end{aligned}$$

The required never married female population at BD is therefore:

$$\begin{aligned}\text{population at GI} + \text{IBDG adjustment factor} &= 95{,}891.5 - 5{,}617\\ &= 90{,}274.5 \text{ never married women}\end{aligned}$$

Having now estimated the three quantities we set out to obtain, we can calculate the required probability that a never married woman turning 20 in 1998 married at age 20:

$$\begin{aligned}\text{probability} &= \frac{\text{marriages of never married women in EBCD}}{\text{never married female population at EB}}\\ &= 11{,}861/96{,}145.4\\ &= 0.12337\end{aligned}$$

and the required first marriage rate at age 20 for never married women who turned 20 in 1998:

$$\begin{aligned}\text{rate} &= \frac{\text{marriages of never married women in EBCD}}{\text{never married female population at BD}} \times 1,000\\ &= (11{,}861/90{,}274.5) \times 1{,}000\\ &= 131.4 \text{ marriages per 1,000 never married women}\end{aligned}$$

Postscript to Example 3

In Chap. 1 we discussed the concept of 'person-years of exposure to risk'. We noted that, strictly speaking, the denominator of a demographic rate should be the number of person-years of exposure to risk for the reference population (the population for which the rate is being calculated) over the reference period (the calendar period, or in the case of a cohort rate, the period of the life cycle for which it is being calculated). However, we discovered that, on the assumption that events subjecting individuals to, and removing them from risk were evenly distributed through the reference period, we could make do with simpler and much more easily obtained approximations (the mean, or mid-period, population at risk).

Similar types of issues arise in relation to the denominators of demographic probabilities. In Example 3 we obtained as the denominator of the probability that a never married woman turning 20 in 1978 married at age 20 the population of never married women at EB (Fig. 3.5); that is, the size of the relevant cohort at the ***beginning*** of the life cycle phase of risk. The definition of the denominator of a probability with which this particular denominator conforms can be stated as: ***The number of person-years that would have been lived between the reference ages/durations by people subject to the risk of the reference event at the first reference age/duration if none of them had experienced the reference event.*** Strictly speaking, this definition is appropriate only when (i) the probability relates to the reference event ***death***, and (ii) one is dealing with a ***closed population***. This is so because:

1. If the reference event is something ***other*** than death, someone who enters the relevant life cycle phase (the reference period) at risk and dies within that phase contributes fully to the denominator of the probability, but cannot contribute to

its numerator during the portion of the reference period that remains after (s)he dies.

2. ***Whatever*** the reference event, an out-migrant during the reference period thereafter can no longer contribute to the numerator of the probability, but contributes fully to the denominator.
3. Similarly an in-migrant during the reference period can contribute to the numerator after arrival, but not having been present at the beginning of the reference period, does not contribute to the denominator.

In practice the ***mortality factor*** (item 1 above) is normally so small that it can be disregarded. With the ***migration factor*** (items 2 and 3) there is obviously some cancelling, in-migrants compensating for out-migrants, so that again it is often possible to disregard the factor for all practical purposes. Only where there is significant ***net*** migration might we have a problem.

If we do decide to make an adjustment for either factor we make it on the assumption that ***each deceased person or migrant was at risk for half the reference period***. Suppose we were to decide to adjust the denominator of the probability calculated in Example 3 for both mortality and migration. Our use of the population of never married women at EB (Fig. 3.5) as this denominator implied two assumptions: (i) that all members of that population thereafter spent a ***full year*** during which they were potentially able to contribute to marriages of never married women aged 20; and (ii) that they were the ***only*** women able to contribute such marriages as members of the cohort in which we were interested. In reality death and out-migration meant that some never married women at EB were able to contribute marriages at age 20 for less than a year, while in-migration meant that newcomers joined our cohort as never married 20 year-olds and were able to contribute marriages for whatever time remained until they reached exact age 21.

In Fig. 3.5 the mortality and migration events we need to consider in making an adjustment are those represented by the parallelogram EBCD. These are the events that occurred to members of our cohort at age 20 (the parallelogram is the intersection of the cohort band and the horizontal band corresponding to age group 20). On the assumption that each in-migration event adds half a person-year of exposure to risk and each death and out-migration event subtracts half a person-year of exposure, our adjustment factor is:

$$\begin{aligned}
&{}^1\!/_2\left[{}^1\!/_2\,(\mathrm{I}_{20,98}+\mathrm{I}_{20,99})-{}^1\!/_2\,(\mathrm{O}_{20,98}+\mathrm{O}_{20,99})-{}^1\!/_2\,(\mathrm{D}_{20,98}+\mathrm{D}_{20,99})\right]\\
&={}^1\!/_2\left[{}^1\!/_2\,(15{,}249+16{,}976)-{}^1\!/_2\,(15{,}090+16{,}432)-{}^1\!/_2\,(52+54)\right]\\
&={}^1\!/_2\,(16{,}112.5-15{,}761-53)\\
&=149\text{ person-years of exposure to risk}
\end{aligned}$$

The adjusted denominator is therefore the sum of the denominator we estimated above and this figure (96,145.4 + 149 = 96,294.4 never married women, or person-years of exposure to risk), and we can use it to calculate an adjusted probability that single women turning 20 in 1978 married at age 20 (11,861 / 96,294.4 = 0.12317). This figure does not greatly alter the result (0.12337) that we obtained ***without*** adjusting, and as this is frequently the case adjustment is often done away with. However, if the added precision is deemed desirable for a particular piece of analysis, adjustment should be undertaken.

Issues Pertaining to Cohort Demographic Processes

From the foregoing discussion of Lexis diagrams and the examples presented illustrating their use you should have gained a sound understanding of the nature of cohort demographic processes. With the cohort approach to demographic analysis we trace the demographic experience of groups of individuals ***as they age together up diagonal bands through a Lexis diagram.*** We turn now to some general issues raised by, or relevant to, this approach.

The Tempo and Intensity of Cohort Processes

The ***tempo*** of a demographic process refers to ***its distribution in age-time or duration-time***. You can think of 'age-time' or 'duration-time' as being a cohort's diagonal band through the Lexis grid. Tempo refers to the distribution of a process up that band. How rapidly or slowly, or how early or late, does it occur through the life course? ***Variations*** in tempo (between two or more populations at a point in time or within a population over time) are variations in the age or duration distribution of a process.

The ***intensity*** of a cohort process refers to ***its level***. The concept can be applied overall (i.e., throughout the life cycle), or with reference to defined phases of the life cycle. The overall intensity of the process of mortality is identical for every population, since everyone eventually dies and does so only once. Other processes, however, involve experiences that are not inevitable in the lives of individuals (e.g., first marriage) or that may be either avoided or repeated in the lives of individuals (e.g., giving birth, or fertility), so that their overall intensity varies between populations and over time for individual populations.

With the process of mortality, identical overall levels of intensity do not preclude variable levels of intensity over defined life cycle phases, although because the overall intensity of the process is known and has a value of unity, or 100 % (death is a once-only, universal experience), such measures are effectively also measures of tempo (of the amounts of a known total occurring within defined life cycle phases). This, however, is a peculiarity of the process of mortality alone.

There are various ways of measuring the intensity of a demographic process. Consider Fig. 3.6, in which a cohort has been defined and e_0, e_1, e_2, etc. refer to

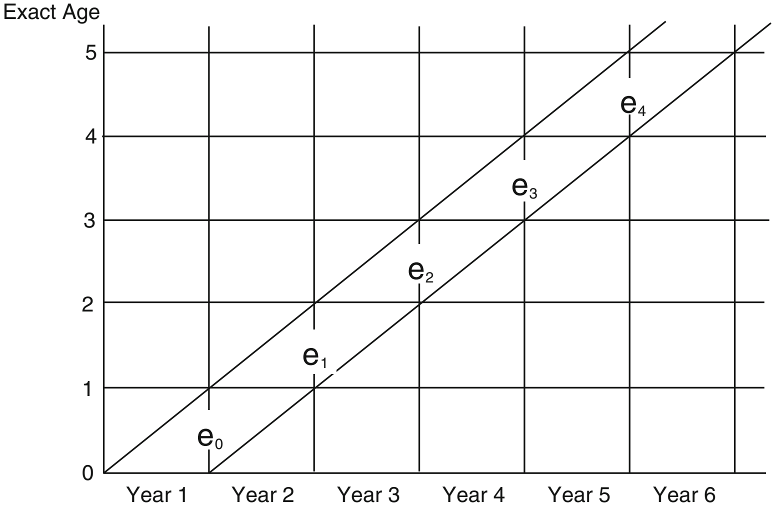


Fig. 3.6 Lexis diagram illustrating the concepts of the 'Intensity' and 'Tempo' of a cohort demographic process

numbers of events of a particular type taking place within successive age-time (or duration-time) parallelograms. We could measure the intensity of this process by:

$$\text{Intensity} = \Sigma_{i=0,\omega}\,(e_i/N_0) \tag{3.1}$$

Where e_i = the number of events occurring at age (or duration) i; N_0 = the population attaining exact age (or duration) 0 (i.e., the original size of the cohort); ω = some age (or duration) beyond which nobody lives.

In the case of mortality, assuming a closed population, this measure is exactly equal to 1 for every human cohort. However, if we let e_i be ***the number of person-years lived*** by the cohort at each age, the measure gives the cohort's expectation of life at birth – a measure of the intensity of ***surviving*** rather than of dying. For the process ***of women giving birth to female children*** the equivalent measure is the cohort ***net reproduction rate***; a measure of the extent to which a population of women replaces itself in the next generation. Another frequently encountered measure of intensity is:

$$\text{Intensity} = \Sigma_{i=0,\omega}\,(e_i/{}_1P_i) \tag{3.2}$$

Where e_i = the number of events occurring at age (or duration) i; ${}_1P_i$ = the number of person-years lived by the cohort between exact ages (or durations) i and i + 1 (approximately the mean population at age (or duration) i); ω = some age (or duration) beyond which nobody lives.

For the process of ***women giving birth to female children*** this measure gives the cohort ***gross reproduction rate***. For the process of ***women giving birth to children*** (of either sex) it gives the cohort ***total fertility rate***. For the process of ***(men or women) marrying for the first time*** it gives the (male or female) cohort ***total first marriage rate***.

A frequently used measure of the tempo of a demographic process takes the general form:

$$\text{Tempo} = \left(\Sigma_{i=0,\omega}\left(i + {}^{1}\!/_{2}\right) \,.\, e_i\right) / \left(\Sigma_{i=0,\omega}\, e_i\right) \tag{3.3}$$

Where e_i = the number of events occurring at age (or duration) i; ω = some age (or duration) beyond which nobody lives.

Equation 3.3 gives the ***mean age*** at which the events of interest in a process occurred (e.g., the cohort mean age at childbearing for the process of women giving birth to children; the male or female cohort mean age at first marriage for the process of men or women marrying for the first time). ***Medians***, together with other similar measures (e.g., quartiles) and, less commonly, ***modes*** are also used as measures of tempo. Medians are often recommended over means where the distribution of an event in age-time or duration-time is heavily ***skewed*** (i.e., instead of approximating the bell-shaped normal curve familiar from statistics, the peak of the distribution is displaced decisively to the right or left, destroying symmetry about that peak). A case in point is the process of first marriage, for which the distribution of events (first marriages) by age often has a pronounced peak quite early in the range of marriageable ages and a long 'tail' extending over older marriageable ages, although this pattern has become less pronounced than formerly in many developed populations as consensual partnering and later marriage have become more common during the 'second demographic transition' (van de Kaa 1987).

Renewable and Non-renewable Events

The distinction between renewable and non-renewable events was drawn in passing in Chap. 1. To recapitulate, a ***non-renewable event*** is one that an individual can experience only once (such as death, first marriage or having a first birth); a ***renewable event*** is one that can recur in an individual's life (such as giving birth, getting married or divorcing).

Renewable events introduce a complexity into the construction and interpretation of demographic measures. They are ***not well suited to the calculation of probabilistic types of measures***. It does not, for instance, make much sense to talk about the probability of giving birth between two exact ages x and x + n, because individual women can contribute more than one event (birth) over this n-year interval. With a probability we aim to measure the likelihood of an event occurring ***once***. Certain specialized types of probabilities can be calculated for renewable event processes;

for example, we could calculate the probability of a woman giving birth ***at least once*** between exact ages x and x + n. This, however, would require detailed data on births by age of mother, birth order and birth interval, and because it eliminated second and subsequent births to the same woman in the age interval would not be a particularly good measure of fertility anyway. Demographic ***rates*** do not pose the same difficulties for measuring renewable event processes, and are therefore the more commonly used type of basic summary measure with such processes.

When it is desired to undertake detailed analysis of renewable event processes, ***a common ploy is to split them into series of non-renewable processes***. Thus, the fertility process can be split into separate processes of first birth, second birth, third birth, and so on. The marriage, or nuptiality, process can be split into processes of first marriage, second marriage, etc., and if we want to recognize the different modes of marriage dissolution that intervene between these processes and their possible implications for remarriage prospects we might think in terms of processes of first marriage, dissolution of first marriage by widowhood and by divorce, second marriage preceded by widowhood and by divorce, dissolution of second marriage by widowhood and by divorce, etc. Splits of this kind pave the way for construction of meaningful probability measures.

For example, if in Eq. 3.1 e_i is the number of ***first*** births occurring at age i, the equation gives, for a closed population, the probability of a woman giving birth at least once, or in other words the probability of becoming a mother. This measure is also known as the cohort lifetime ***parity progression ratio*** from parity 0 to parity 1. The concept of parity, or the number of live births a woman has had, has been touched on previously, and parity progression will be dealt with more fully in Chap. 6 when the analysis of fertility is discussed. But in similar fashion probabilities of progressing between any other pair of successive parities can be computed. In general, any cohort lifetime parity progression ratio beyond parity 1 is given by:

$$\mathrm{PPR} = \Sigma_{i=y,\omega}\, b_{i,k+1} / \Sigma_{i=y,\omega}\, b_{i,k} \tag{3.4}$$

Where k and k + 1 are the two parities between which progression is being measured; $k \neq 0$; b = live births; i = a single-year age group; y, ω = respectively the youngest and oldest ages at which any woman gives birth.

Equation 3.4 is the ratio of the total number of parity k + 1 births to the total number of parity k births for a cohort throughout its life. It can also be applied to progression from parity 0 to parity 1 by making the denominator N_0 (the original size of the female cohort), since all women attain parity 0.

Attrition

Attrition is ***the process by which a risk population is depleted*** (i.e., reduced in size) ***over time as it experiences a non-renewable event***. In a non-renewable demographic process, once a person has experienced the reference event (s)he ceases to be at risk of experiencing it and is no longer part of the population under analysis. Through age-time or duration-time the population at risk therefore gets progressively smaller. It is to this process of getting smaller that the word ‘attrition’ refers.

Demographers often study attrition processes, or non-renewable event processes, by means of ***attrition tables***, or ***life tables***. Strictly speaking a life table is an attrition table which summarizes the non-renewable process of death, or mortality. However, the term ‘life table’ has acquired a more general usage, such that these days any attrition table is called a life table, and indeed the latter term is far more frequently used than the former.

Attrition tables are constructed from probabilities of experiencing the reference event between successive exact ages or durations (normally at single-year intervals, although for a process like the termination of breastfeeding following birth, intervals of a month would be more appropriate while mortality attrition is often summarized using 5-year intervals). These tables typically feature the following columns:

1. The number of reference events occurring in each age or duration group, or interval. When constructing an attrition table for a cohort, this column, obtained empirically, along with knowledge of the initial size of the cohort (N_0 in the notation introduced with Eq. 3.1) is often the basis from which the rest of the table is constructed. In the conventional, or standard, life table (summarizing the non-renewable event process of death), values in this column are the ${}_1d_x$ values – ${}_1d_0$, ${}_1d_1$, ${}_1d_2$, etc. (see Chap. 4).
2. The population at the start of each age or duration interval; that is, the initial population at risk at exact age/duration 0, and thereafter the survivors from that population yet to experience the reference event. The term ‘survivors’ is obviously especially apt when the reference event is death, and remains quite appropriate when it is something like divorce, but it borders on comical when the reference event is, say, first marriage (‘survivors in the never married state’). In the notation introduced with Eq. 3.1 these values are denoted by N_0, N_1, N_2, etc., while in the notation of a conventional life table they are the l_x values (l_0, l_1, l_2, etc.). If N_0 and ${}_1d_x$ values have been obtained empirically, remaining N_x values can be found using the relationship $N_x = N_{x-1} - {}_1d_{x-1}$.
3. The age-specific or duration-specific probabilities of experiencing the reference event within each age or duration interval. Denoted in the conventional life table by ${}_1q_x$ (${}_1q_0$, ${}_1q_1$, ${}_1q_2$, etc.), this column can be obtained using the relationship ${}_1q_x = {}_1d_x / N_x$ (or ${}_1d_x / l_x$). However, as we shall see when the conventional life table is discussed more fully in Chap. 4, it may also be derived independently by other means and become the key to constructing the remainder of the attrition table.
4. The number of person-time units (person-years, person-months, etc., depending on the age/duration interval employed in the table) spent at risk (or ‘lived’) by survivors from the initial population at risk within each age/duration interval. In conventional life table notation this is the ${}_1L_x$ column, and provided a uniform distribution of reference events by exact age/duration within an age/duration

interval can be assumed, these values can be obtained from the relationship ${}_1L_x = (N_x + N_{x+1})/2$ (or $(l_x + l_{x+1})/2$).

5. Cumulated person-time units spent at risk above each exact age/duration. These are the T_x values in conventional life table notation, and are obtained by summing all values of ${}_1L_i$ where $i \geq x$.
6. The mean number of person-time units (person-years, person-months, etc.) remaining before the event occurs per 'survivor' at each exact age/duration. This is the $e^o{}_x$, or 'expectation of life', column of the conventional life table and is obtained using the equation $e^o{}_x = T_x/N_x$ (or T_x/l_x).
7. It is also not uncommon to find a column of age-specific (or duration-specific) rates of experiencing the reference event. Denoted in the conventional life table by ${}_1m_x$ (${}_1m_0$, ${}_1m_1$, ${}_1m_2$, etc.), this column is generated using the relationship ${}_1m_x = {}_1d_x/{}_1L_x$.

The conventional life table is the most common type of attrition table in demography, but the principles it employs can be used in analysing any other non-renewable demographic process. Table 3.1 is an example of an attrition table for the non-renewable process of termination of breastfeeding after birth of an infant. Breastfeeding patterns are of interest to demographers in their search for understanding of fertility levels and trends in different populations, since breastfeeding acts as a form of contraceptive, delaying the resumption of ovulation after birth of a child. Aside from any nutritional arguments, the encouragement of breastfeeding is often advocated to lengthen birth intervals, and thus lower fertility, in high fertility populations.

The input data in Table 3.1 are the ${}_1d_x$ values together with N_0, which were obtained from a survey. The table relates to second-last births of women (to avoid the complication of women still breastfeeding their youngest child at interview), and excludes a few cases where the child was not breastfed or died before breastfeeding ended. There are clear questions over the quality of the data, with obvious tendencies for respondents to have opted for breastfeeding durations of 12 and 18 months over durations either side of those ones (see ${}_1d_x$ column). The degree to which this reflects them having set those durations as targets in advance or having approximated their durations of breastfeeding in retrospect is unknown. Note also that the table has been truncated at $x = 24$ months, so that (i) the last row deals with terminations of breastfeeding at durations of 24 months or longer ('24+' months), and (ii) since all women who reached exact duration 24 months still breastfeeding had to terminate at some duration thereafter, the probability of termination in this 'open' duration interval (q_{24+} – the first subscript is omitted because the interval is not of width 1 month) is 1.0000. Columns of the table apart from ${}_1d_x$ were generated using relationships specified above. The final value in the ${}_1L_x$ column assumes that all women still breastfeeding at exact duration 24 months terminated breastfeeding during the twenty-fifth month after on average breastfeeding for half that month.

The 'attrition' in Table 3.1 is evident in the N_x column, where the initial 2,164 breastfeeding women are depleted month by month until after 24 months only 32 are still breastfeeding. The $e^\circ{}_x$ column gives average periods of breastfeeding remaining

Table 3.1 Breastfeeding attrition table for Sudan, 1979

Month (x)	$_1d_x$	N_X	$_1q_x$	$_1L_x$	T_x	e°_x
0	4	2,164	0.0018	2,162	15,700	7.3
1	27	2,160	0.0125	2,147	13,538	6.3
2	79	2,133	0.0370	2,094	11,392	5.3
3	150	2,054	0.0730	1,979	9,298	4.5
4	337	1,904	0.1770	1,736	7,319	3.8
5	274	1,567	0.1749	1,430	5,584	3.6
6	323	1,293	0.2498	1,132	4,154	3.2
7	389	970	0.4010	776	3,022	3.1
8	142	581	0.2444	510	2,247	3.9
9	111	439	0.2528	384	1,737	4.0
10	72	328	0.2195	292	1,353	4.1
11	10	256	0.0391	251	1,061	4.1
12	154	246	0.6260	169	810	3.3
13	6	92	0.0652	89	641	7.0
14	6	86	0.0698	83	552	6.4
15	6	80	0.0750	77	469	5.9
16	6	74	0.0811	71	392	5.3
17	1	68	0.0147	68	321	4.7
18	25	67	0.3731	55	254	3.8
19	2	42	0.0476	41	199	4.7
20	2	40	0.0500	39	158	4.0
21	3	38	0.0789	37	119	3.1
22	2	35	0.0571	34	83	2.4
23	1	33	0.0303	33	49	1.5
24+	32	32	1.0000	16	16	0.5

Source: Sudan Fertility Survey, 1979

for women still breastfeeding at each exact duration of breastfeeding since birth. Thus the first value in it tells us that the 2,164 women on average breastfed for 7.3 months. This figure then generally declines, but rises sharply again at exact duration 13 months – i.e., immediately after the attrition of women who claimed to have breastfed for 12 months. This indicates that women continuing to breastfeed beyond that point were very committed to doing so. Those still breastfeeding after 13 months on average continued to do so for a further 7.0 months, so on average breastfed for 20.0 months.

Period Analysis and Synthetic Cohorts

We began this chapter by noting that demographic analysis adopted either of two approaches: the cohort approach or the period approach. Thus far we have concentrated exclusively on the cohort approach, which in many ways is the preferred analytic perspective in demography. It has ***a realism*** that derives from the fact that it focuses on an actual group of individuals and traces their experience through the life course. Moreover, ***most demographic measures are based in some way or other on measures of the intensity or tempo of demographic processes for cohorts***.

The cohort approach does, however, have drawbacks***. Demographic data are usually not collected on a cohort basis.*** Registration data tend to be tabulated annually, the events covered corresponding to vertical columns in the Lexis grid. Similarly census and survey data usually refer to a time instant (the latter, perhaps, in an approximate sense since interviewing may take several days or weeks), and therefore correspond to a vertical line through the Lexis grid. Surveys which collect retrospective life histories take a cohort approach to the gathering of data. But cohort analysis (running diagonally through the Lexis grid) frequently ***draws on many separate annual series of data***, adding to its complexity. More crucially***, cohorts take time to live out their lives***, and we cannot afford to wait around until they have finished their lives before starting to study demographic processes affecting them. We can, and do, study the ***partial***, or ***incomplete***, experience of cohorts, up to whatever stages of their lives they have reached at the time of analysis. This does not necessarily obscure demographic change as it occurs, but it can be an inefficient way of demonstrating it.

Given these drawbacks, demographers often engage in ***period analysis***. This is a ***cross-sectional*** approach to analysis in which the focus is either on events occurring in a particular year (or consecutive group of years) or on events occurring to a population at a particular point in time in the age/duration groups its members occupied at that time. Thinking in terms of the Lexis diagram, we saw that cohort analysis is concerned with events located along ***diagonal bands*** drawn through the Lexis grid. ***Period analysis is concerned with events arranged VERTICALLY through the Lexis grid*** (Fig. 3.7). Either they are located in columns of the Lexis grid (analysis is focused on events that occurred in a particular year or series of consecutive years – the upper diagram in Fig. 3.7), or they are located in a series of parallelograms stacked vertically through the grid about a line representing a focal population (analysis is concerned with events that occurred to members of this population in the age/duration groups they occupied at the focal time, each group's experience being represented by its relevant parallelogram of cohort experience – the lower diagram in Fig. 3.7).

It is common with period analysis to treat events occurring in either of the two Lexis configurations just described ***AS IF they were experienced by a cohort***. This practice gives rise to a distinction between ***real cohorts*** and ***synthetic cohorts***. Real cohorts are the 'genuine' cohorts we have been dealing with thus far. They are the ones represented by diagonal bands through a Lexis diagram. Synthetic cohorts, by comparison, are ***hypothetical*** cohorts, represented on a Lexis diagram by either of the configurations illustrated in Fig. 3.7 and, as those two diagrams illustrate, ***composed of segments of the experience of a series of real cohorts***. You can see in Fig. 3.7 that diagonal bands representing several successive real cohorts have been drawn in, and that the shaded synthetic cohorts corresponding to demographic

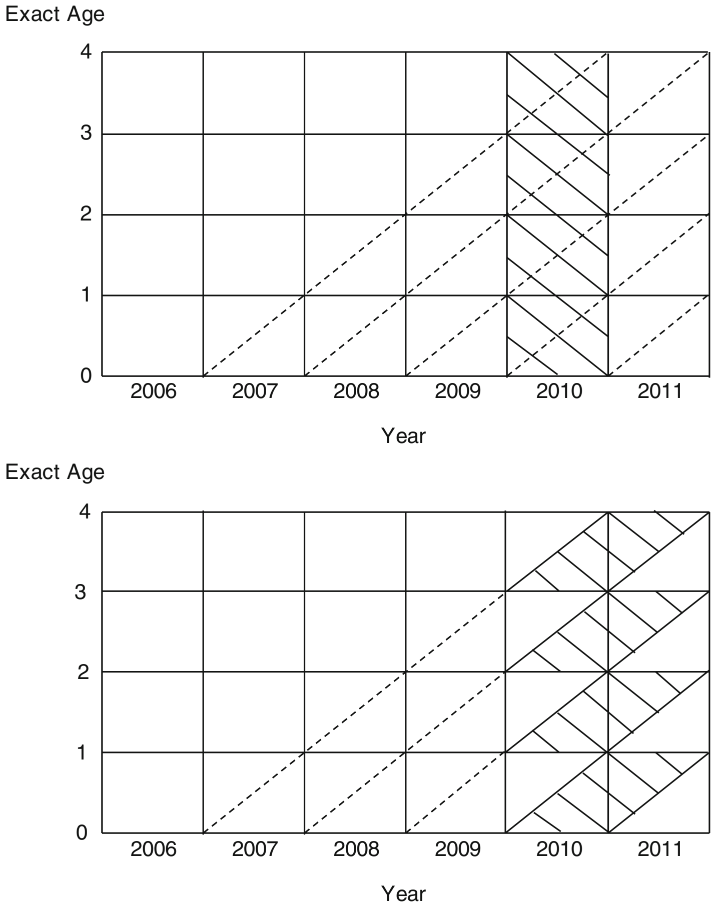


Fig. 3.7 Alternate Lexis diagram representations of a synthetic cohort

experience during calendar year 2010 (upper diagram) and of the population at the end of 2010 (lower diagram) comprise little bits (segments) of the experience of each of these real cohorts. Thus the calendar year 2010 synthetic cohort consists of:

Half the experience at age (or duration) 0 of the 2010 real cohort
Half the experience at each of ages (or durations) 0 and 1 of the 2009 real cohort
Half the experience at each of ages (or durations) 1 and 2 of the 2008 real cohort
Half the experience at each of ages (or durations) 2 and 3 of the 2007 real cohort
Etc.

Similarly the 31 December 2010 synthetic cohort consists of:

The experience at age (or duration) 0 of the 2010 real cohort
The experience at age (or duration) 1 of the 2009 real cohort

The experience at age (or duration) 2 of the 2008 real cohort
The experience at age (or duration) 3 of the 2007 real cohort
Etc.

Note that the little bits of real cohort experience that we cobble together to form synthetic cohorts ***collectively cover all possible life course phases, without duplicating any phase***. Note also that when period analysis is concerned with measures constructed for synthetic cohorts it essentially asks the following question:

> What would be the life cycle experience of this demographic event of a real cohort which happened to encounter the age-specific (or duration-specific) rates or probabilities of experiencing the event which prevailed (i) in this particular year, or (ii) cross-sectionally among the population at this particular date?

This is a ‘what if?’ question. ***What*** would a real cohort’s life cycle experience be ***if*** it happened to experience this cross-sectional regime of age-specific (or duration-specific) rates or probabilities. You can perhaps now appreciate why synthetic cohorts are sometimes called ***hypothetical*** cohorts.

The Tendency of Synthetic Cohort Measures to Exaggerate Change

Period measures that purport to measure the lifetime experience of cohorts (i.e., that measure the lifetime experience of ***synthetic*** cohorts) ***are apt to take on values that are abnormally high or low***. If a trend line plotting such a measure over time is compared with one plotting the equivalent measure for successive ***real*** cohorts, typically the line for the synthetic cohort measure has more pronounced peaks and troughs than that for the real cohort measure.

Figure 3.8 illustrates this phenomenon. It shows for New Zealand over a lengthy period the total fertility rate (TFR) and the cohort completed fertility rate (CCFR). The TFR is a synthetic cohort measure which, we noted in Chap. 1, is found by summing age-specific fertility rates in a year over all reproductive ages, and which indicates the number of children women would on average have during their reproductive lives ***assuming*** they experienced that year’s schedule of age-specific fertility rates (in other words it gives the average completed family size for the synthetic cohort). The CCFR gives the ***actual*** average completed family size for (real) birth cohorts of women. It is plotted in Fig. 3.8 with birth cohorts offset 27 years against calendar year synthetic cohorts, age 27 being an approximate mean age of childbearing for the period covered. Thus the CCFR plot at calendar year 1921 on the horizontal axis is the plot for the 1894 birth cohort (the cohort that turned 27 in 1921). For the birth cohorts of 1962–1971, CCFRs (plotted against calendar years 1989–1998) are estimates based on recorded fertility to at least age 40 and then, increasingly as cohorts become more recent, projections of future fertility at ages 40–49. Because fertility is relatively low at ages 40–49 these projections

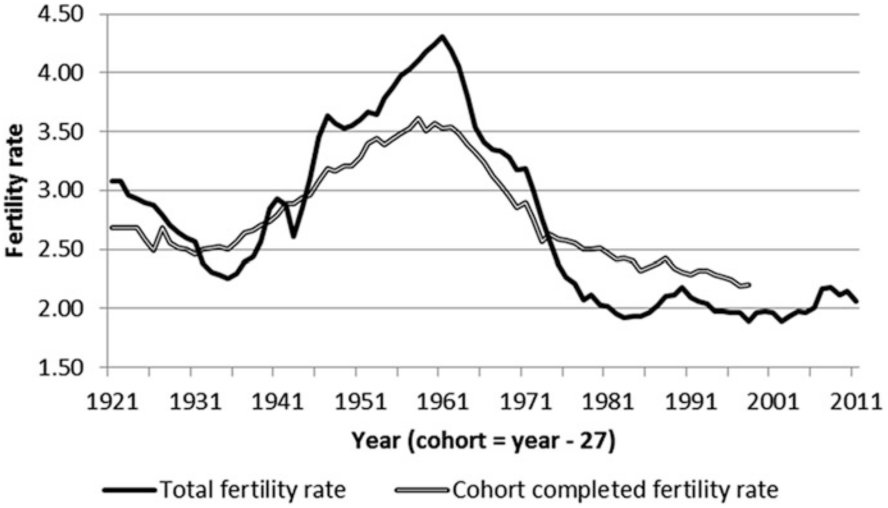


Fig. 3.8 Trends in New Zealand's total fertility rate and cohort completed fertility rate (Source: Statistics New Zealand. Note: Cohort completed fertility rates cover 1894–1971 birth cohorts)

will almost certainly ultimately prove to be quite accurate, and allow the CCFR trend line to be extended a further 10 years. It is not, however, extended beyond the 1971 birth cohort (plotted against calendar year 1998) because the birth cohorts in question were in 2011 (the final year for which the TFR is plotted) still aged less than 40. This, of course, serves to illustrate one of the drawbacks of cohort analysis noted above – the availability of only partial data for some cohorts.

You can see in Fig. 3.8, however, that the TFR 'troughs' in the mid-1930s and briefly in the early 1940s, then again in the early 1980s at levels lower than the CCFR for birth cohorts then of peak childbearing age (i.e., cohorts of the early 1900s and late 1950s respectively). Similarly it 'peaks' in 1960 at over 4.2 children per woman, well above the CCFR peak of 3.6 children per woman recorded for the 1931 birth cohort. A difference of 0.6 of a child per woman in estimates of the peak average family size associated with the post-war baby boom may not seem large, but it is a substantial one.

So ***what causes this tendency for synthetic cohort measures like the TFR to exaggerate the extent of demographic change?*** We noted above that synthetic cohorts add together bits, or segments, of the experience of series of real cohorts, and one answer to the question is that, at times, they ***add together particularly favourable or unfavourable segments***. This happens when there is cross-sectional ***heaping*** of demographic events, or when cross-sectional ***deficits*** of events occur. In the former circumstance cross-sectional age-specific or duration-specific rates or probabilities of event occurrence are generally higher than any real cohort contributing part of the synthetic cohort will experience across its life; in the latter

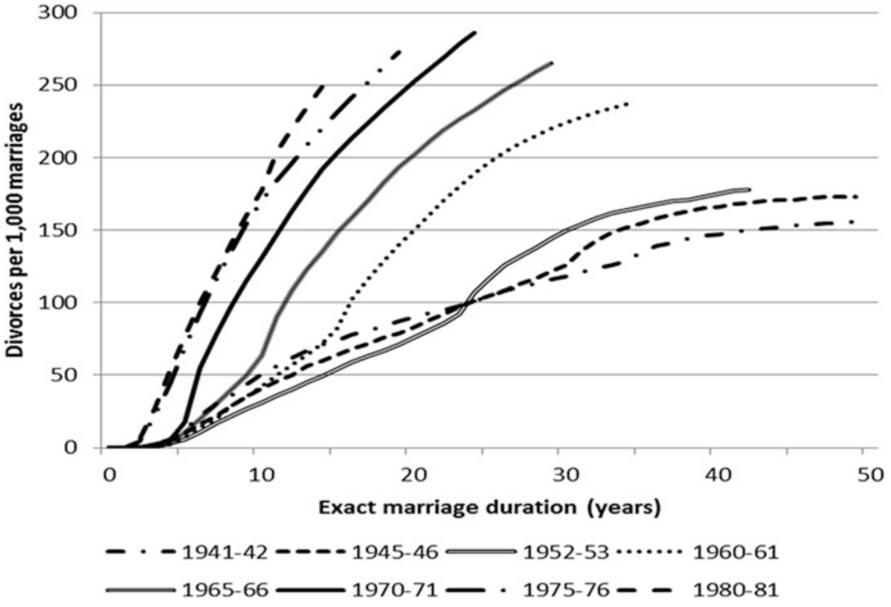


Fig. 3.9 Cumulative divorce trajectories to 1994 for selected Australian marriage cohorts (Source: Adapted from Carmichael et al. (1997: Figure 4))

circumstance they are generally lower. Two mechanisms, either separately or in combination, can bring these circumstances about.

1. ***Demographic behaviour often responds to historical events or periods which are either especially conducive or especially unfavourable to that behaviour.*** For example, marriage activity and childbearing may be reduced in times of war or severe economic depression, and conversely may be strongly stimulated by the cessation of war or economic recovery. The Great Depression and the outbreak of the Second World War clearly had the former effect in New Zealand (Fig. 3.8). Liberalizing divorce laws similarly invariably gives rise to an immediate, across-the-board increase in synthetic cohort measures of the incidence of divorce, these implying a level of instability of marriage higher than is characteristic of any real marriage cohort. These sorts of ***'period effects'***, as they are known, tend to affect entire populations when they occur; different real cohorts are affected simultaneously at whatever stages of the life cycle they happen to have reached at the time, leading to a sudden surge or decline in the cross-sectional incidence of the demographic behaviour in question.

   Figure 3.9 illustrates this. It shows for selected financial year (July–June) Australian marriage cohorts, followed until 1994, cumulative divorce rates to successive exact marriage durations. Each cohort trajectory at one point steepens sharply, this steepening occurring at successively earlier marriage durations for successively more recent marriage cohorts. What is in evidence

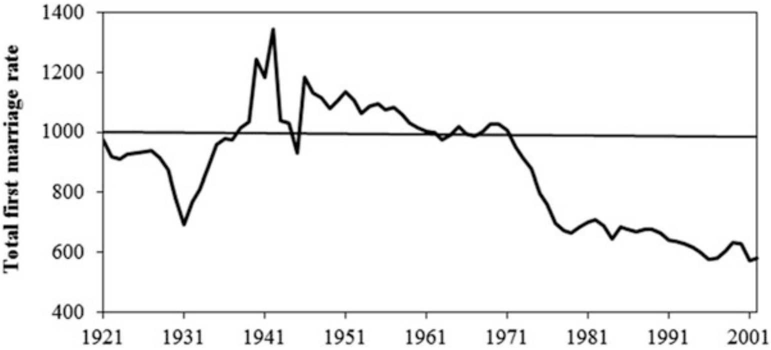


Fig. 3.10 Total first marriage rate for Australian females, 1921–2002 (Source: Adapted from Carmichael (2002: Figure 1))

is the introduction in 1976 of Australia's *Family Law Act*, which substantially liberalized the grounds for divorce. There was a rapid, in part temporary, surge in the number of divorces granted (i) as couples who would have had to wait to divorce under the old law became immediately eligible to do so under the new one, and (ii) as people took the new law to signal greater public acceptance of dissolving unhappy marriages. Marriage duration-specific divorce rates recorded in 1976 implied that, in a synthetic marriage cohort constructed for that year, 62 % and almost 71 % of marriages would end in divorce within 30 years and 50 years respectively. But these were one-off figures, unlikely to be remotely approached by any real marriage cohort in the foreseeable future. The equivalent 1977 synthetic cohort figures were much lower at 44 and 50 %, and by 1979 they were down to 36 and 40 % as backlogs of disrupted marriages at all marriage durations were cleared.

When a synthetic cohort measure is calculated for a period or a time instant that is especially auspicious or inauspicious for the demographic behaviour being studied, the inference is that a real cohort could sustain the abnormally high or low ***cross-sectional*** incidence of the behaviour prevailing at that time throughout an actual life cycle. In reality this is rarely the case. ***Real cohorts tend to compensate for abnormally high or low levels of experience of a demographic event at one stage of the life cycle by lowering or raising their levels of experience at later stages.*** Figures 3.10 and 3.11 are again illustrative. The former shows the total first marriage rate (TFMR) for Australian females since 1921, a synthetic cohort measure analogous to the TFR which indicates the proportion of a birth cohort who would ever marry if it experienced through life the age-specific ratios of first marriages to total population recorded in the year for which the calculation is performed. The feature of this graph to note at this point is the huge trough centred on 1931. This reflects the economic disincentive to marriage generated by the Great Depression.

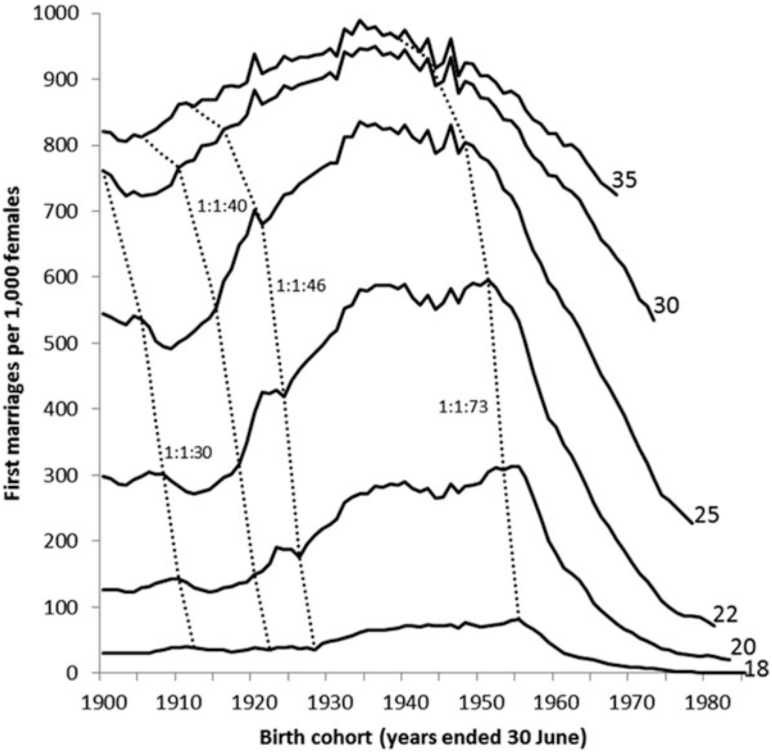


Fig. 3.11 Cumulative first marriage rates to selected exact ages for Australian female birth cohorts (Source: Adapted from Carmichael (1988: Figure 4))

If we then focus on Fig. 3.11 we find proportions of successive female ***birth cohorts*** in Australia ever married by exact ages 18, 20, 22, 25, 30 and 35 plotted. The effect of the Depression is evident in 'dips' in trend lines between time lines drawn at 1:1:30 and 1:1:40. But if we note the birth cohorts for which the dips are evident at exact ages 20 and 22 (those of 1911–1919 and 1909–1917 respectively), then examine the trend for the same cohorts at exact age 35, the dip has disappeared. By this life cycle stage these cohorts had not only made up for reduced marriage activity in their teens and early twenties; they'd become integrated into a trend toward more universal marriage.

2. The second mechanism that causes synthetic cohort measures to exaggerate the extent of change over time is ***change in the tempo, or timing, of a demographic process***; that is, a tendency, over time, for the process to occur earlier or later in the life cycle. Consider Fig. 3.12, on which is represented (i) the first marriage experience of a 1990 synthetic cohort of women attaining marriageable age in Australia (exact age 16) (the shaded vertical column), and (ii) the first marriage experience of a series of real cohorts of women attaining marriageable age between 1976 and 1990, each of which contributed part of their experience to

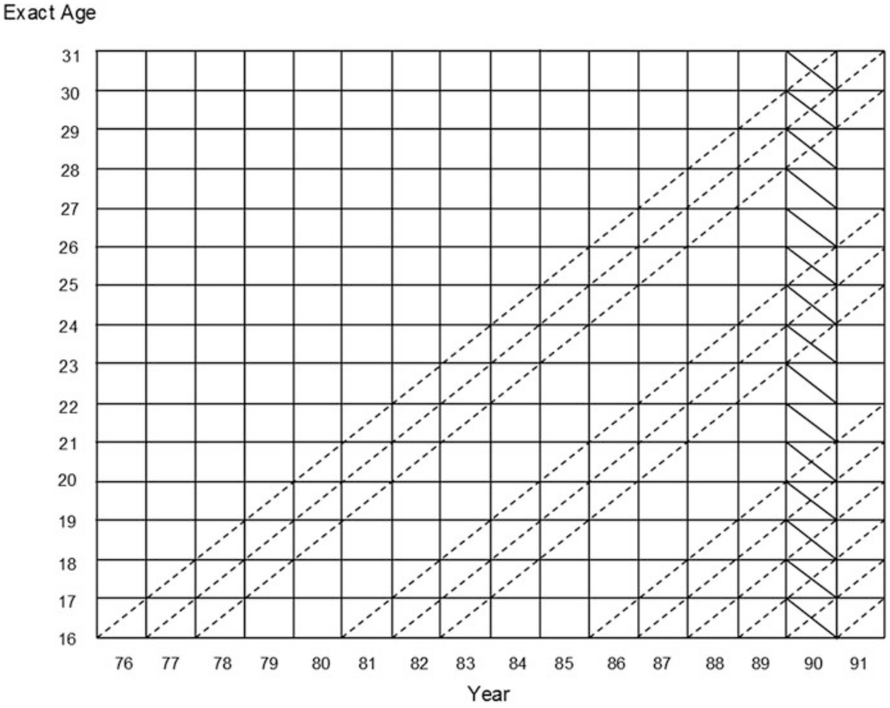


Fig. 3.12 Lexis diagram illustrating how a change in the tempo of the process of first marriage leads to abnormally high or low numbers of first marriages occurring cross-sectionally in synthetic cohorts

the synthetic cohort (the dashed diagonal bands – note that before 1986 only selected bands are shown, but you can imagine where others would run). The overlap between the diagonal bands and the shaded vertical column defines the parts of the experience of the various real cohorts that are added together to create the synthetic cohort.

Suppose that a trend to earlier marriage was in progress in this population around 1990. This means that, in that year, the number of first marriages at ***younger*** ages will be ***relatively high*** (compared to recent years), because real cohorts that reached marriageable age more recently (those of 1986–1990, say) will be leading the trend to earlier marriage (and such a trend implies proportionately more people marrying at younger ages). However**,** assuming that the propensity to ever marry remains fairly constant, the number of first marriages at ***older*** marrying ages in 1990 will ***also*** be ***relatively high*** (compared to what women in real cohorts supplying marriages at younger ages to the 1990 synthetic cohort will experience 10 years or so into the future when ***they*** reach the older marrying ages). Why? We have to think of what the real cohorts supplying first marriages to our synthetic cohort at older ages (late 20s and early 30s) experienced when ***they*** were younger. Figure 3.12 indicates that the cohorts concerned reached marriageable age 10–15 years previously, in the mid to late 1970s. Since we have a trend to earlier marriage in progress we can assume that these cohorts married ***less*** freely when they were young than the real cohorts of 1986–1990. It follows that when these earlier cohorts reached the older ages at which they were contributing to the 1990 synthetic cohort, ***relatively large numbers of their members were still unmarried*** and at risk of marrying for the first time at those older ages (compared to the situation that would pertain in the future when the 1986–1990 real cohorts reached their late 20s and early 30s). Thus, the 1990 synthetic cohort combines the experience of younger real cohorts whose first marriage experience is relatively heavily weighted toward ***younger*** ages with that of older real cohorts whose experience is relatively heavily weighted toward ***older*** ages. The result is an abnormally high number of female first marriages occurring ***cross-sectionally*** in 1990. We call this phenomenon cross-sectional ***heaping***.

Alternatively, suppose that a trend to later marriage is in progress around 1990. This means that, in that year, the number of first marriages at ***younger*** ages will be ***relatively low*** (compared to recent years), because real cohorts that reached marriageable age more recently will be leading the trend to later marriage (and such a trend implies proportionately fewer marriages at younger ages). However, the number of first marriages at ***older*** marrying ages in 1990 will also be ***relatively low*** (compared to what women in real cohorts supplying first marriages at younger ages to the 1990 synthetic cohort will experience 10 years or so into the future when ***they*** reach the older marrying ages). Why? Again we have to think of what the real cohorts supplying first marriages to our synthetic cohort at ages in the late 20s and early 30s experienced when younger. Since a trend to later marriage is in progress we know that these cohorts married ***more*** freely when they were young than the real cohorts of 1986–1990 (Fig. 3.12). It follows that when they reached the older ages at which they were contributing marriages to the 1990 synthetic cohort, ***relatively small numbers of their members were still unmarried*** and at risk of marrying for the first time at those older ages (again, compared to the situation that would pertain in the future when the 1986–1990 real cohorts reached their late 20s and early 30s). Thus, the 1990 synthetic cohort combines the experience of younger real cohorts whose first marriage experience is ***relatively lightly weighted toward younger ages*** with that of older real cohorts whose experience is ***relatively lightly weighted toward older ages***. The result is an abnormally ***low*** number of female first marriages occurring cross-sectionally in 1990, a phenomenon we refer to as a cross-sectional ***deficit***.

The translations between real and synthetic cohort experience which accompany changes in tempo and have been outlined here for the process of first marriage apply equally in respect of other demographic processes. They can be complicated if a change in intensity accompanies the change in tempo. But ignoring this possibility and maintaining the focus on the process of first marriage, the capacity of changes in tempo to cause synthetic cohort measures to assume exaggeratedly high or low values is amply demonstrated by Fig. 3.10.

This graph features a horizontal line at TFMR = 1,000. This is the value of the TFMR that implies universal marriage, and clearly over long periods the TFMR for Australian females ***exceeded*** that level. That this is possible is at first perplexing, but all a TFMR above 1,000 means is that first marriages have heaped cross-sectionally in the year in question at a level that is ***unsustainable in the longer term***. Somewhere in the recent past, or down the track, there has been, or will be, a compensating deficit of first marriages.

The first period during which the TFMR exceeds 1,000 coincides with the Second World War, at which time extreme cross-sectional heaping of first marriages was due to the circumstances of war (i.e., the 'historical events/periods' mechanism was operating). There was a 'now or never' rush to the altar (or registry office) by couples seeking to marry before grooms departed for war service overseas, and this was aided by the fact that, with separation imminent, the usual economic prerequisites to marriage could be largely disregarded. Then, later during the war, first marriage activity received a considerable boost from the presence of American servicemen in Australia on leave. It is estimated that there were 12,000–15,000 marriages of American servicemen to Australian women during the War, much having been written about the attraction their uniforms, their wallets, their readiness for a good time and their capacity to provide gifts of merchandise that was in short supply (e.g., nylon stockings) represented (Aitchison 1972; Moore 1981; Sturma 1989).

The second period during which the female TFMR exceeded 1,000 lasted from the end of the war (when the return of servicemen temporarily created another special stimulus to marriage) to the early 1960s. This sustained period of high TFMRs, which really lasted through to the early 1970s, was in large part the product of declining ages at first marriage, which were an integral part of Australia's post-war marriage boom and therefore its baby boom. In other words it was largely a product of the 'changing tempo' mechanism outlined above. Later on, through the 1970s, the female TFMR fell steeply, until by the end of the decade it was at a level that suggested that only about two-thirds of a cohort attaining marriageable age would ever marry. The evidence for real cohorts is that this figure is exaggeratedly low; that perhaps 20–25 %, rather than a third, will fail to marry. Again the changing tempo mechanism was primarily responsible for this exaggeration, because commencing in the early 1970s the previous trend to earlier marriage reversed emphatically and was replaced, as part of the second demographic transition, by a strong trend to later marriage.

Other Issues in Period Analysis

We often make assumptions and use approximations in constructing period measures. This is especially so when period analysis follows the model illustrated by the upper diagram in Fig. 3.7. Figure 3.13 illustrates the translation of real cohort concepts in the construction of demographic measures using this model. In a real

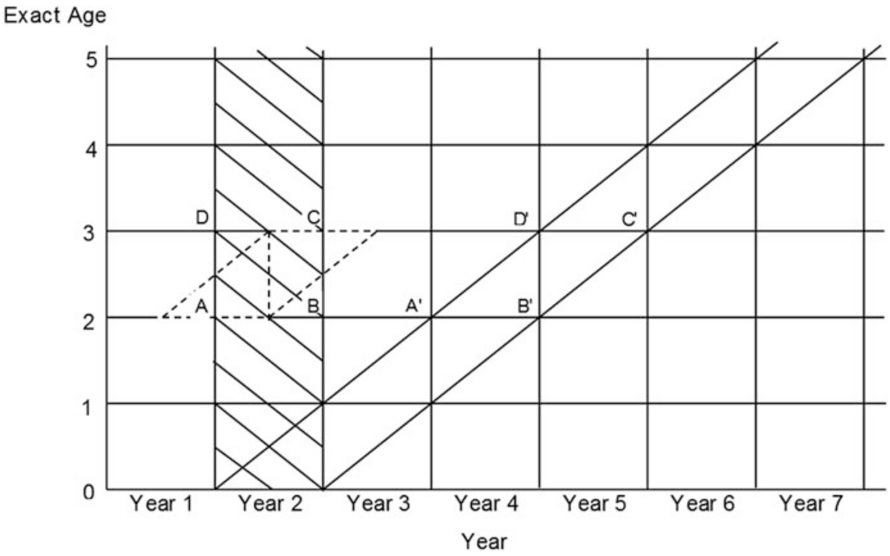


Fig. 3.13 Lexis diagram illustrating the translation between real cohort and synthetic cohort demographic probabilities and rates

cohort the events considered in constructing a rate or probability for any age or duration group lie in a parallelogram; for example, in the parallelogram A′B′C′D′. In a period synthetic cohort of the type depicted in the upper diagram of Fig. 3.7 they lie in a square (e.g., ABCD). In the real cohort case, the events in A′B′C′D′ all occur to the population at A′B′; that is, to the population attaining the exact age/duration marking the start of the relevant age or duration interval. However, in the period synthetic cohort case events in the triangle ACD ***do not*** occur to members of the population at AB. The period equivalent of the cohort ***probability*** *events in A′B′C′D′ divided by the population at A′B′*, that is, *events in ABCD divided by the population at AB*, assumes that events in ACD give a reasonable estimate of events in the equivalent triangle whose left side is the line BC. Similarly the period equivalent of the cohort ***rate*** *events in A′B′C′D′ divided by the population at B′D′*, that is, *events in ABCD divided by the population at a line joining the midpoints of AB and DC* (the dashed vertical line), is perhaps best thought of as approximating the rate for the cohort represented by this line. The relevant events for the ‘true’ rate for this cohort are contained not within ABCD, but within the dashed parallelogram centred on the dashed vertical line. This parallelogram overlaps substantially with the square ABCD, but its left and right extremities extend into the grid squares immediately adjacent to ABCD on its left and right. Our period rate ***essentially assumes that the number of events lying inside ABCD but outside the dashed parallelogram is a reasonable approximation of the number lying inside the parallelogram but outside ABCD***.

All measures used in cohort analysis have equivalents in period analysis, but many summary measures which have simple constructions as cohort measures must be calculated in different ways as period measures so as to make allowance for the differences between period and cohort populations. Measures such as the net reproduction rate and expectation of life, which are simple measures of intensity for a cohort, become more complicated when calculated from period data.

Period measures should not, however, be seen as second-best to cohort measures, despite their derivative status. Many influences on demographic processes, as we have already seen, are time-specific rather than specific to the life experience of particular cohorts. The impact of a famine, a war, an influenza pandemic, the introduction of a new treatment for a particular disease, a major change in divorce legislation, etc. will be felt by all age/duration cohorts in a population at the time it occurs, not by a particular cohort or group of cohorts. Similarly, where demographic processes are affected by economic conditions or changes in social attitudes these effects also tend to be time-specific rather than cohort-specific (although there may be variability across cohorts in a matter such as revision of social attitudes – younger cohorts are often more receptive to social change than are older ones). There is a good case to be made that most influences on demographic processes are period, or cross-sectional, or time-specific influences, and this means that a need for period measures would exist even if cohort measures were available in a timely fashion (which they are not). As already noted, cohort analysis smoothes out sudden peaks and troughs in period measures that are caused by transient circumstances. It tends to focus on permanent, as opposed to transient, change. But demographers are interested in both types of change, and so there is a place for both the period and the cohort approaches to demographic analysis.

There are cases where the calculation process that gives meaningful results for a cohort gives very misleading results when applied to period analysis. For example, if we have cohort data on numbers of children ever born classified by mothers' ages, it is possible to calculate age-specific fertility rates for the cohort of women from the differences between mean numbers of children ever born to the cohort at different ages. However, because fertility is a very volatile demographic variable the same technique applied cross-sectionally to a population, seeking to combine the experience of different cohorts of women with possibly very different fertility histories, is usually spectacularly unsuccessful.

With any type of period analysis, the interpretation of measures is helped by considering the principles underlying the construction of equivalent measures in the cohort case. Nearly all measures have very simple interpretations for cohorts, which may be difficult to discern for equivalent period measures.

References

Aitchison, R. (1972). *Thanks to the Yanks? The Americans and Australia*. Melbourne: Sun Books.

Carmichael, G.A. (1988). *With this ring: First marriage patterns, trends and prospects in Australia*. Canberra: Department of Demography, Australian National University and Australian Institute of Family Studies.

Carmichael, G. A. (2002). Taking stock at the millennium: Family formation in Australia. In G. A. Carmichael & A. Dharmalingam (Eds.), *Populations of New Zealand and Australia at the millennium* (pp. 91–104). Canberra/Wellington: Australian Population Association/Population Association of New Zealand.

Carmichael, G. A., Webster, A., & McDonald, P. (1997). Divorce Australian style: A demographic analysis. In C. A. Everett (Ed.), *Divorce and remarriage: International studies* (pp. 3–37). New York: The Haworth Press.

Moore, J. H. (1981). *Over-sexed, over-paid and over here: Americans in Australia 1941–1945*. St Lucia: University of Queensland Press.

Sturma, M. (1989). Loving the alien: The underside of relations between American servicemen and Australian women in Queensland, 1942–1945. *Journal of Australian Studies, 13*(24), 3–17.

Van de Kaa, D. (1987). Europe's second demographic transition. *Population Bulletin*, *42*(1), 1–57.

Chapter 4
Analysis of Mortality: The Life Table and Survival

Introductory Matters

In earlier chapters we have played around a bit with the crude death rate (CDR). We have seen that, while it is the most readily available measure of mortality, true to its name it is crude, to the point of being at times extremely misleading. Indeed, other than as a component of population growth, it is a measure to be avoided.

In our efforts to sort out a perplexing differential between CDRs for Australia and Malaysia we discovered that any meaningful analysis of mortality levels and patterns must pay heed to the extreme variability in the incidence of death by age in particular, and also by sex. We looked initially at calculating death rates specific for age and sex, and followed up by making use of such rates in applying techniques of standardization and decomposition to expose the capacity for serious distortion in a simple comparison of CDRs. Age and sex are not the only dimensions of population composition along which mortality varies; as was noted in Chap. 2, mortality levels in some populations vary across such additional dimensions as ethnicity, education level, occupation and marital status. In Australia, for instance, there are major differences between the mortality patterns of the Indigenous and non-Indigenous populations; mortality can be higher among the least well educated sector of a population because of poorer living conditions, more limited knowledge of behaviours and environmental conditions that endanger health, and more limited awareness of and access to modern medical care; some occupations (e.g., mining, building and construction) are inherently more hazardous than others are; and being unmarried sometimes carries a higher mortality risk, in part because poor health selects certain people out of marriage. But variation in mortality levels by age and sex is universal.

One consequence of this is that ***the analysis of mortality typically deals with patterns for males and females separately***. In this chapter we will examine the life table, the major tool of the demographer in the analysis of general mortality

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_4

patterns and trends. Life tables are usually constructed separately for males and females. Concerning differentials in the level of mortality by age, probably the most important is the invariably higher level of mortality in infancy (the first year following live birth) than at other childhood ages. ***Infant mortality is an especially important category of mortality***. It is both a major component of total mortality in many (particularly less developed) populations, and by far the category of mortality where the largest number of ***avoidable*** deaths occurs. Thus, the greatest potential for reducing ***overall*** mortality often rests with reducing ***infant*** mortality. Moreover, in some populations a reduction in infant mortality is doubly desirable in that added certainty over the survival of children is likely to be a force for fertility decline. Infant mortality is also important because, since they occur at the very beginning of life, infant deaths averted have the greatest potential for improving overall survivorship in a population as measured by the well known summary measure ***average expectation of life at birth***. In many ways the infant mortality rate is itself a summary measure of the overall level of mortality in a population; high mortality and low mortality populations typically have high and low infant mortality rates respectively. Infant mortality is also central to the study of ***child survival***, a prominent specialist field of policy-relevant research within demography. First, though, it is necessary to deal with general measures and methods in mortality analysis.

We need to be clear at the outset just what it is we are studying when we study mortality. ***What is 'death'?*** The World Health Organization (1977: 15) defines it as 'the permanent disappearance of all evidence of life at any time after live birth has taken place.' The main point to note about this definition is its specific exclusion of ***foetal deaths*** (deaths 'prior to the complete expulsion or extraction from its mother of a product of conception ... [such that] after ... separation the foetus does not breathe or show any other evidence of life' (World Health Organization 1977: 13) – in other words, miscarriages, abortions and stillbirths). The deaths in which demographers are interested for the purpose of studying general mortality patterns and trends presuppose 'live birth' (the definition of which will be left until the analysis of fertility is discussed).

The Life Table: A General Perspective

As noted above, the life table is the major tool available to demographers for the general analysis of mortality. A life table is ***constructed from age-sex-specific death rates***. While life tables typically are regarded as being concerned with mortality, they should more properly be thought of as being concerned with ***survival***; that is, with ***not*** dying rather than with dying. Viewed this way a life table becomes a tool for focusing on ***the health of a population***, as indexed by ***the quantity of life*** its members enjoy. There are also, of course, issues of ***quality*** of life. In its basic form the life table does not directly address these issues, although quantity and quality are to a considerable degree correlated, and the life table approach can be extended to encompass issues of quality of survival.

The life table is also appropriately thought of as a device concerned with survival in the context of population projections and associated planning applications. Demographers with planning responsibilities often need to estimate numbers of people who will occupy the age groups for which particular types of facilities and services will be needed at nominated dates in the future, and estimating survivors from a census population and/or projected future birth cohorts who will form the relevant age groups at the relevant dates is, together with a consideration of the likely impact of migration, fundamental to such an exercise. The life table provides us with ***a theoretical framework from which to calculate survivors***.

A life table is an attrition table that traces the process of mortality (or survival). In its usual ***period*** form it is a good example of the application of cohort concepts to period data. It is obviously impractical to wait around for over 100 years until a birth cohort dies out before constructing a complete ‘cohort’ record of mortality. Cohort life tables ***can*** be constructed for ***partial*** cohorts, but they are partial life tables in two senses. Not only do they pertain to partial cohorts; they do not include some of the more important life table functions (in particular the ‘average expectation of life remaining’ function), which cannot be calculated until a cohort’s complete mortality experience is known. Given these difficulties, conventional practice is to use cross-sectional data to construct life tables which trace the mortality experience of ***synthetic*** birth cohorts. These ***product-limit*** life tables, as they are known because of their distinctive method of construction, address the following question:

> ***What would be the lifelong process of mortality attrition of a birth cohort of men or women which experienced the age-specific death rates that prevailed for the relevant sex in year y?***

An obvious further question arises. Is it reasonable to assume that persons born in year y will, 60 and more years into the future, say, experience the year y death rates at ages 60 and over? People forming these age groups in year y might have been exposed to risk factors injurious to their health earlier in their lives that younger people may be less affected by; for example, younger people might be more aware of the health risks associated with smoking than were older people when they were younger, and may therefore smoke less. Medical advances might in the future assist inherently more frail individuals to survive to older ages, reducing death rates at younger ages but increasing them at older ages. These are very plausible propositions, but without a crystal ball, ***can we do better than construct our life tables using the latest information available?*** This seems to satisfy life insurance companies, whose premiums are based on detailed analyses of life tables. They probably console themselves by assuming that most future changes in mortality regimes are likely to be improvements, and therefore to their advantage, although the emergence during the 1980s of HIV-AIDS warned against complacency on that front. Then again, it is not as if, for populations covered by the insurance industry, new life tables are not produced at regular intervals, so that mortality conditions at different ages are updated (and premiums adjusted) as any real cohort approaches those ages.

There are two types of period life table: ***single-year-of-age*** life tables (also known as ***standard*** or ***complete*** life tables) which, as their name indicates, treat each year of life separately; and ***abridged*** life tables, which are based on death rates for fewer, broader age intervals. While other schemes of abridgement are possible, abridged life tables typically use death rates for 5-year age groups, except that (i) the first year of life is treated separately, so that age group 0–4 is split into age groups 0 and 1–4, and (ii) the oldest ages (frequently ages 85 and over) are grouped into a single category.

The Single-Year-of-Age Life Table

Although excessively detailed for many demographic purposes, the single-year-of-age life table is the more straightforward variant when it comes to understanding the mechanics of life table construction. It begins with a group of people (of a given sex) at birth and follows them through life, subtracting each year the number who die. We have already seen, in the example using Sudanese breastfeeding data presented in the previous chapter, that the principles of attrition can be readily extended to the study of non-renewable demographic processes other than mortality, but restricting ourselves to mortality, the sorts of questions a single-year-of-age life table permits us to answer include:

1. What is the probability of surviving (or dying) between two nominated birthdays, from one single-year age group to some older single-year age group, or from a nominated birthday to an older single-year age group?
2. What is the average life expectancy of men, or women, at birth, or at any subsequent birthday? Life expectancy at birth is a fairly familiar concept given the frequency with which it is used to index general population health and improvements or deterioration therein, but one can also, using a life table, calculate the average ***remaining*** expectation of life given survival to any nominated birthday.
3. What would be the age structure of a ***stationary population*** experiencing constantly the mortality regime represented by the life table? A stationary population is one that experiences a constant annual number of births, a constant mortality regime (i.e., a constant set of age-specific death rates) that annually produces exactly the same number of deaths as births, and is closed to migration. Such a population has a constant size, a zero growth rate and a constant age structure. We will return to stationary populations when addressing stable population theory. In the meantime it is sufficient to note that what we call a ***life table population*** is a stationary population. We will return to the idea of a life table population shortly.

Obtaining Age-Sex-Specific Death Rates to Construct a Life Table

The age-sex-specific death rates from which we construct a single-year-of-age life table are given by:

$$ {}_1M_x = ({}_1D_x / {}_1P_x)\,1{,}000 \tag{4.1}$$

Where ${}_1M_x$ denotes the death rate for males or females aged x last birthday (i.e., between exact ages x and x + 1); ${}_1D_x$ = male or female deaths at age x during the year for which the calculation is being performed (i.e., deaths between exact ages x and x + 1); ${}_1P_x$ = the mid-year male or female population aged x last birthday (i.e., the population aged between exact age x and exact age x + 1).

We noted in Chap. 1 that more exact denominator terms for demographic rates, expressed in terms of person-years of exposure to risk, are possible. However, unless extreme precision is required, and unless one's data are of sufficient quality for the necessary calculations to be performed, a denominator more sophisticated than the one used in Eq. 4.1 is rarely justified.

What ***may*** be justified is an adjustment to the numerator term of Eq. 4.1, in recognition of the fact that relatively small numbers of deaths of persons of a given sex occur annually in some single-year age groups. This can introduce an element of ***random fluctuation*** into single-year-of-age life tables, and as a means of smoothing out such fluctuations a common practice is to use, not deaths at age x in the year for which a life table is being constructed, but the annual ***average*** of deaths at age x ***over a period of years*** centred on the year of interest. Thus in Australia and New Zealand, for example, new life tables are prepared after each census using not just deaths in the census year, but averages of deaths over a 3-year period straddling the census year (i.e., for the census year and one year either side of it). In this scheme the equation equivalent to Eq. 4.1 becomes:

$$ {}_1M_x = \left\{\left[{}^1\!/_3\left({}_1D_x^{(y-1)} + {}_1D_x^{(y)} + {}_1D_x^{(y+1)}\right)\right] / {}_1P_x^{(y)}\right\} 1{,}000 \tag{4.2}$$

Where y denotes the year on which the life table is to be centred; other items have the same meanings as in Eq. 4.1.

For actuarial purposes (i.e., when life tables are constructed for use by, for example, life insurance companies) much more smoothing of data than Eq. 4.2 accomplishes typically is engaged in, with denominators as well as numerators of age-sex-specific death rates likely to receive attention. The aim of actuaries is to generate a smooth underlying set of age-specific mortality probabilities that eliminates irregularities reflecting temporary or chance factors. Demographers, however, are much less worried about these irregularities. Indeed, they are even somewhat suspicious of the lengths to which actuaries go to eliminate them, because sometimes their sources lie in demographically interesting phenomena.

Components of a Life Table

A life table begins with ***an arbitrary number of individuals at birth*** (i.e., at exact age 0). This arbitrary number is known as the ***radix*** of the life table. In theory it could have any value one might care to give it, but for the sake of convenience a number which is some power of 10 is almost always used, and by far the most common radix to encounter is 100,000.

Having selected a radix (and if you don't choose 100,000 you are probably being awkward), a single-year-of-age life table is constructed by applying probabilities of dying between successive birthdays initially to the radix population, and thereafter to survivors from that population until the point is reached at which there are no longer any survivors. Thus we start by applying the probability of dying between birth (exact age 0) and exact age 1 to the radix population and calculate survivors at exact age 1. We then apply the probability of dying between exact age 1 and exact age 2 to survivors at exact age 1 and obtain survivors at exact age 2. And so on through each successive single year of age.

Life tables consist of a series of columns of numbers, each column giving values of a particular ***life table function***. These columns are related to one another by a series of equations and can be derived from one another. Indeed, much of the information in a life table can be said to be redundant; different columns are equivalent, and merely say the same thing in different ways. While you will encounter different numbers of columns in life tables, there is a basic core of six life table functions that are found in most single-year-of-age life tables. They have a standard, internationally recognized notation, and you should familiarize yourself with that notation and with what the various life table functions measure. Some of these notations were introduced in passing when we addressed the general concept of 'attrition' in Chap. 3.

The six core life table functions are:

$_1\mathbf{q}_\mathbf{x}$ – the probability of dying between exact age x and exact age x + 1 (or if you like, the proportion of people who reach exact age x alive who die before reaching exact age x + 1).

$\mathbf{l}_\mathbf{x}$ – the number of members of the initial radix population surviving at exact age x (the radix itself is denoted by l_0). (Note that in the breastfeeding attrition table for Sudan presented as Table 3.1 in Chap. 3 the equivalent function was denoted by N_x, but in a mortality attrition table (i.e., a 'true' life table) l_x is the conventional notation.)

$_1\mathbf{d}_\mathbf{x}$ – the number of members of the initial radix population who die between exact ages x and x + 1.

$_1\mathbf{L}_\mathbf{x}$ – the number of person-years lived between exact ages x and x + 1 by members of the initial radix population (or, if you like, by those of them who survive to exact age x).

$\mathbf{T}_\mathbf{x}$ – the number of person-years lived at ***all*** ages ***above*** exact age x by members of the initial radix population (or, if you like, by those of them who survive to exact age x).

e^o_x – the average number of years of life remaining beyond exact age x for each member of the initial radix population who survives to exact age x.

Other life table functions that you may encounter include:

$_1m_x$ – the life table death rate between exact ages x and x + 1. (Note that this life table death rate should be distinguished from $_1M_x$, which is the ***observed*** age-specific death rate between exact ages x and x + 1. The two are not necessarily equal in some methods of life table construction, although methods described in this chapter ***assume*** equality. Be aware, however, of the ***conceptual*** difference between them. One ($_1m_x$) is a life table function; the other ($_1M_x$) is obtained not by manipulating other life table functions, but from empirical data – see Eqs. 4.1 and 4.2 above.)

μ_x – the instantaneous force of mortality at exact age x.

$_1p_x$ – the probability of surviving between exact age x and exact age x + 1 (or if you like, the proportion of people who reach exact age x alive who survive to reach exact age x + 1).

$_1S_x$ – the survival ratio; the proportion of people who survive to the age group bounded by exact ages x and x + 1 from the immediately younger single-year age group, or from birth to the very youngest age group (that bounded by exact ages 0 and 1).

Note in the above list of functions that some are prefixed by a subscript 1, and others are not. Functions that ***do*** have a prefixed subscript 1 are quantities that pertain to a 1-year ***interval*** of age – the interval stretching from exact age x to exact age x + 1; functions that ***do not*** have a prefixed subscript 1 are quantities that pertain either to ***exact*** age x (l_x, μ_x and e^o_x are in this category) or to ***all ages beyond*** exact age x (as with T_x). Having said this, however, it is not uncommon for single-year-of-age life table functions which strictly speaking ***should*** be written with a prefixed subscript 1 to in fact be written ***without*** it. Thus you may encounter q_x instead of $_1q_x$, d_x instead of $_1d_x$, etc. Don't be confused by this. If the prefixed subscript is missing, you can assume it to be 1, and that you are dealing with a life table function for a single-year-of-age life table. The prefixed subscript is often dropped as an unnecessary piece of detail when its value is invariably 1 (i.e., when it is clear that the life table in question is a single-year-of-age life table).

Generating a Single-Year-of-Age Life Table: The $_1q_x$ Column

The normal starting point when constructing a single-year-of-age, or product-limit, life table is the $_1q_x$ column. This is the column we derive from raw age-specific data on deaths and population size. There are a number of approaches to the task. Which one we use depends to a degree on the nature of the input data available and the ages for which calculations are to be performed. We will begin with a simple estimator of $_1q_x$ in a closed population (one whose size is not affected by migration) which is

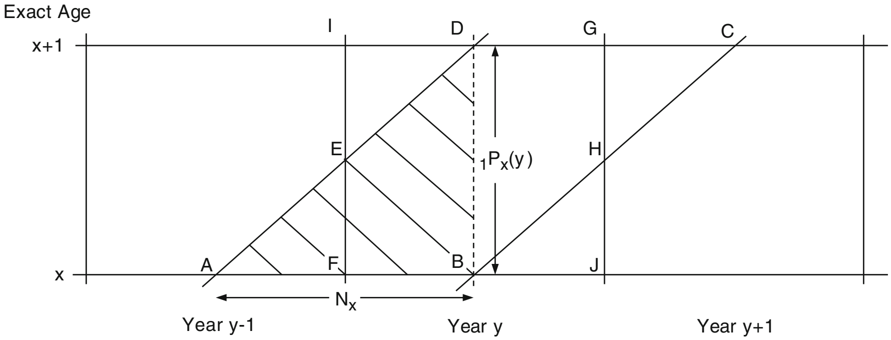


Fig. 4.1 Lexis diagram illustrating derivation of standard equation for obtaining values of $_1q_x$

applicable provided $x \neq 0$ (***the calculation of $_1q_0$ is invariably treated as a special case*** for reasons that will be discussed shortly).

Consider a closed population at the middle of some reference year y. The probability of dying at age x for the population aged x in the middle of year y (which we'll denote by $_1P_x^{(y)}$) is given by:

$$_1q_x = {}_1D_x/N_x \tag{4.3}$$

Where $_1D_x$ = deaths at age x to the population aged x in the middle of the reference year y (i.e., to the population $_1P_x^{(y)}$ – note that the 'D' is italicized to distinguish it from the one that appears in Eqs. 4.1 and 4.2); N_x = the size of this cohort at exact age x (i.e., at the beginning of the life cycle phase during which its members were at risk of dying at age x).

The Lexis diagram presented as Fig. 4.1 will assist us to evaluate this equation, because almost certainly we will not have directly available to us data to substitute for $_1D_x$ and N_x in Eq. 4.3. The line BD in Fig. 4.1 represents the population aged x in the middle of year y (i.e., $_1P_x^{(y)}$); it is the intersection of the vertical line representing the population in the middle of year y and the horizontal band corresponding to age group x (the age group bounded by *exact* ages x and x + 1). It is this population, or cohort, for which we wish to obtain a value for $_1q_x$.

We next draw in the diagonal lines passing through AD and BC which define, or enclose, the demographic experience of our cohort. The deaths at age x experienced by this cohort ($_1D_x$) are those that lie within the parallelogram ABCD; this parallelogram is the intersection of the diagonal band representing the cohort, and the horizontal band corresponding to age group x. Data on deaths available to us will be tabulated by single years of age and year of occurrence; i.e., they will correspond to ***squares*** of the Lexis grid. Clearly the parallelogram ABCD lies partly within three different Lexis grid squares (it is the sum of the triangle AFE, the area EFBHGD and the triangle GHC), so that applying principles learned in Chap. 3:

$$ {}_1D_x = {}^1\!/_8 D_x^{(y-1)} + {}^3\!/_4 D_x^{(y)} + {}^1\!/_8 D_x^{(y+1)} \tag{4.4} $$

Where $D_x^{(y)}$ = deaths at age x during year y (the equivalent of ${}_1D_x^{(y)}$ in Eq. 4.2 above).

In conformity with our definition of a probability from Chap. 1, the denominator N_x of the righthand side of Eq. 4.3 is the size of the cohort of interest at exact age x, which marks the ***beginning*** of the life cycle phase when a person is at risk of dying aged x. Thus N_x is represented in Fig. 4.1 by the line AB; the intersection of the diagonal band representing the cohort and the horizontal line representing exact age x. How do we estimate N_x? Well, again applying principles learned in Chap. 3, the size of our cohort at AB is its size at BD (for which we are likely to have data) ***plus*** the deaths represented by the shaded triangle ABD. This triangle captures the demographic experience (in this case the mortality) of the cohort between exact age x and the middle of year y, during which time mortality was ***reducing*** its size. But we are working ***backwards*** from the middle of year y to exact age x (an ***earlier*** point in the cohort's life course), and so we have to ***add back in*** the deaths represented by the shaded triangle. This triangle lies partly in each of two squares of the Lexis grid (the triangle AFE plus the area EFBD), and so applying standard Lexis diagram principles we have:

$$ N_x = {}_1P_x^{(y)} + {}^1\!/_8 D_x^{(y-1)} + {}^3\!/_8 D_x^{(y)} \tag{4.5} $$

We have now obtained expressions for both elements in the righthand side of Eq. 4.3 in terms of the size of the population aged x in the middle of year y (${}_1P_x{}^{(y)}$) and numbers of deaths at age x in specified calendar years ($D_x^{(y)}$). These are quantities for which data are typically available, and so we can proceed to calculate a value for ${}_1q_x$.

The equation expressing ${}_1q_x$ as the ratio of the righthand sides of Eqs. 4.4 and 4.5 is, however, cumbersome, and a simplified form is often used. This rests on the assumptions, first, that deaths in the triangle AFE in Fig. 4.1 are reasonably approximated by deaths in the triangle EDI, and second, that deaths in the triangle GHC are reasonably approximated by deaths in the triangle BJH. Under these assumptions Eqs. 4.4 and 4.5 reduce to:

$$ {}_1D_x = D_x^{(y)} $$

and:

$$ N_x = {}_1P_x^{(y)} + {}^1\!/_2 D_x^{(y)} $$

The equation for ${}_1q_x$ thus becomes:

$$ {}_1q_x = D_x^{(y)} / \left({}_1P_x^{(y)} + {}^1\!/_2 D_x^{(y)} \right) $$

or, simplifying by dropping superfluous superscripts and subscripts:

$$q_x = D_x / \left(P_x + {}^1\!/_2 D_x \right) \tag{4.6}$$

Where D_x = deaths at age x during the year for which q_x is required; P_x is the mid-year population aged x in that year.

It is also possible to express Eq. 4.6 in terms of the observed age-sex-specific death rate at age x, which we denoted in Eqs. 4.1 and 4.2 by ${}_1M_x$. If we define this death rate by ${}_1M_x = M_x = D_x / P_x$ (i.e., don't multiply by 1,000 as in Eq. 4.1), we can rearrange to obtain $D_x = M_x.P_x$. Substituting for D_x in Eq. 4.6 yields:

$$q_x = M_x \,.\, P_x / \left(P_x + {}^1\!/_2 M_x \,.\, P_x \right)$$

whence, cancelling:

$$q_x = M_x / \left(1 + {}^1\!/_2 M_x \right) \tag{4.7}$$

or:

$$q_x = 2M_x / \left(2 + M_x \right) \tag{4.8}$$

Equation 4.7 or 4.8 (they are equivalent) is commonly used to obtain values of q_x from age-sex-specific death rates when $x \neq 0$. Remember, though, that if you have death rates ***per 1,000*** persons at risk, ***you will need to divide them by 1,000*** (i.e., shift the decimal point three places to the left) before substituting in Eq. 4.7 or 4.8.

Note also that Eq. 4.7 or 4.8 yields ***standard approximations*** of q_x values given values of M_x. They are approximations on two counts. First, we derived the equation assuming that we were dealing with a closed population. In fact populations rarely are closed, but departures from such a situation normally are not substantial enough to be of consequence. Second, we introduced assumptions to allow us to base our calculation on deaths in a single calendar year when, strictly speaking, we should have based it on deaths over a 3-year period. Be aware that other, more complex methods exist for obtaining q_x values. They need not concern you now, save to note (i) that they may be called for if precision is required in circumstances of appreciable net migration and/or rapid annual change in mortality levels, and (ii) that their use may be indicated if you are unable to reproduce q_x values in a published life table from raw deaths and population data. Another possible explanation, should you strike the latter situation, is that M_x values were subjected to ***graduation***, or smoothing, before they were used in evaluating Eq. 4.7 or 4.8.

Obtaining $_1q_0$: Separation Factors

But, ***what about $_1q_0$, or q_0***, which we noted needed to be treated as a special case? This is so because the Lexis diagram principles that underpin the derivation of Eqs. 4.6 and 4.7 (or 4.8) are invalid at age 0. They assume ***even*** distributions of events (deaths in this case) within age-time squares of the Lexis grid. As we have already noted, for infant mortality ***this assumption is patently false. Infant deaths are very heavily concentrated, by exact age, at the lower end of the age interval bounded by exact ages 0 and 1***; that is, the great majority of infant deaths occur very shortly after birth.

This means that, whereas for any cohort reaching exact age x ($x \neq 0$) during a reference year approximately half the deaths at age x usually occur in that year and half in the following year, for persons ***born in*** the reference year (i.e., reaching exact age 0) the majority of deaths at age 0 occur ***during the year of birth***. Because of this a satisfactory estimate of q_0 is normally provided by the infant mortality rate (IMR) (remember, it was pointed out in Chap. 1 that the IMR was an estimate of the probability that a person born during the year in question died in infancy; all we are saying now is that, usually, it is a ***pretty good*** estimate). Thus:

$$q_0 = \mathrm{IMR} = D_0/B \tag{4.9}$$

Where $D_0 =$ deaths at age 0 in the reference year y; $B =$ live births during year y.

An estimate of q_0 may also be obtained where we have available data on deaths at age 0 and the mid-year population aged 0, but not data on births. From the former data, using Eq. 4.1 (without the multiplier of 1,000), we can calculate the death rate at age 0, M_0, which we can use in conjunction with a quantity called a ***separation factor*** to compute q_0. The demographic literature provides various ways of calculating this separation factor, but a simple approximation, which conveniently also makes use of M_0, can be obtained using the following formula derived empirically by Keyfitz (1970):

$$f = 0.07 + 1.7M_0 \tag{4.10}$$

q_0 is then given by:

$$q_0 = M_0/\left(1 + (1 - f)\,M_0\right) \tag{4.11}$$

Note, however, that in practice this equation can pose problems because of ***differential under-enumeration at age 0 in registration and census data***. Other equations for q_0 (not least Eq. 4.9) use only registration data (on births and infant deaths). Because it is based on M_0, Eq. 4.11 uses both registration and census data (for the numerator and denominator, respectively, of M_0), and in developed countries under-enumeration at age 0 tends to be more of a problem with census data. Respondents filling out census forms sometimes overlook children aged 0, perhaps because of difficulty understanding the concept of someone being aged ‘0’. Young babies are typically thought of as being x ***days***, ***weeks*** or ***months*** old, where x is a non-zero number, not 0 ***years*** old.

Setting this issue to one side, the obvious question is, ***what is a separation factor?*** Note at the outset, lest any misconception arise due to the concept being dealt with at this juncture, that separation factors are ***not*** only used in deriving the q_x column of a life table, and are ***not*** only used when dealing with infant mortality (i.e., mortality at age 0). Both points should become clear as this chapter proceeds.

For a 1-year birth cohort attaining exact age x, deaths at age x occur partly during the year in which cohort members attain exact age x and partly during the following year. In this single-year-of-age context, ***a separation factor measures the proportion of deaths at age x to a cohort attaining exact age x during year y which occur during year y + 1***. If you compare the righthand sides of Eqs. 4.11 and 4.7 above you will note a similarity. The latter is a generalized version of the former in which the separation factor f takes the value ½, or 0.5 (since $1 - f = ½$, then $f = ½$). A separation factor is therefore implicit in the standard equation for q_x when $x \neq 0$. That its value is 0.5 reflects the equation’s underlying assumption that deaths to the cohort at age x are ***evenly distributed*** through the relevant age-time parallelogram on a Lexis diagram (see Fig. 4.1); in this circumstance half the deaths at age x occur in the year during which the cohort attains exact age x (between the middle of year $y - 1$ and the middle of year y in Fig. 4.1), and half occur during the following year (between the middle of year y and the middle of year $y + 1$).

In the case of infant mortality we have already noted that deaths concentrate heavily at the ***beginning*** of the 1-year interval of exposure to risk (i.e., towards exact age 0 rather than towards exact age 1). The effect of this is, if you like, to distort the cohort diagonals on the Lexis diagram after the fashion illustrated in Fig. 4.2. Deaths at age 0 of members of the year y birth cohort lie within what we might conveniently conceptualize as a ***curved*** band rather than a diagonal band, the

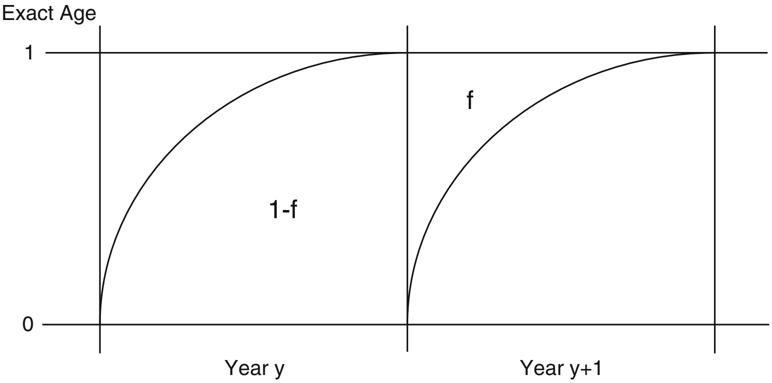


Fig. 4.2 Lexis diagram illustrating the separation factor f used in generating certain life table functions at age 0

majority (a proportion measured by 1 – f) occurring during the year of birth and a minority (a proportion measured by f) occurring during the following year. Thus, the separation factor f that we ***might*** use in calculating q_0 (depending on which equation for q_0 we use), and ***will*** use in calculating other life table functions pertaining to the first year of life, ***always has a value less than 0.5***. This is true regardless of which of the variety of available approaches to obtaining a value for f is used.

Note, though, that this is a characteristic of ***this particular*** separation factor (i.e., that used in calculating life table functions pertaining to the first year of life). Separation factors are also used at older ages, especially in constructing abridged life tables where, because we are dealing with age intervals wider than 1 year (typically 5 years), the assumption that deaths are distributed evenly through an interval by exact age can also be untenable. Many of these separation factors take on values ***greater*** than 0.5 (but less than 1.0). In general, ***if the tendency is for mortality to concentrate at the YOUNGER end of an age interval a separation factor for the age interval will have a value LESS THAN 0.5; if the tendency is for mortality to concentrate at the OLDER end of an age interval a separation factor for the age interval will lie BETWEEN 0.5 AND 1.0***. This statement should suggest to you another general interpretation for separation factors. ***A separation factor measures the proportion of deaths in an age interval that occur in the second half of the age interval***. If deaths are evenly distributed through the age interval by exact age the separation factor has a value of 0.5, or one-half. If they are concentrated early in the interval or late in the interval the separation factor is respectively lower than, or greater than, this value (with the proviso it can never lie outside the range 0.0–1.0).

Having moved from a discussion of methods of estimating q_0 to one of separation factors, we will now briefly backtrack and note that we can use the separation factor f to modify our original equation for q_0, Eq. 4.9. This sets q_0 equal to the infant mortality rate, the ratio of deaths at age 0 in the reference year y to the number of live births in year y. But as we noted previously, and as Fig. 4.3 illustrates, some of the deaths in the numerator of this ratio (a proportion of them equal to the separation factor f) involved children born not in year y, but in year y – 1 (the previous year) (this is a slightly different interpretation of f to that illustrated in Fig. 4.2). Because our deaths derive partly from live births in year y – 1 and partly from live births in year y, there is a case for respecifying Eq. 4.9 so that its denominator, instead of being live births in year y, becomes ***a weighted average of live births in years y – 1 and y, with the separation factor f providing the weight***. Hence we have:

$$q_0 = D_0 / \left(fB^{y-1} + (1 - f)\, B^{y} \right) \tag{4.12}$$

When annual numbers of births are reasonably constant from year to year there is no need to bother going to the trouble of using Eq. 4.12. The simpler Eq. 4.9 is adequate. However, ***where the sizes of successive birth cohorts differ substantially*** it is preferable to use Eq. 4.12.

Note in passing that where Eq. 4.12 or Eq. 4.9 is used in preference to Eq. 4.11 to obtain q_0, care should be taken that the value of m_0, the life table death rate at age 0, is compatible with q_0. It was noted above that problems can arise with Eq. 4.11

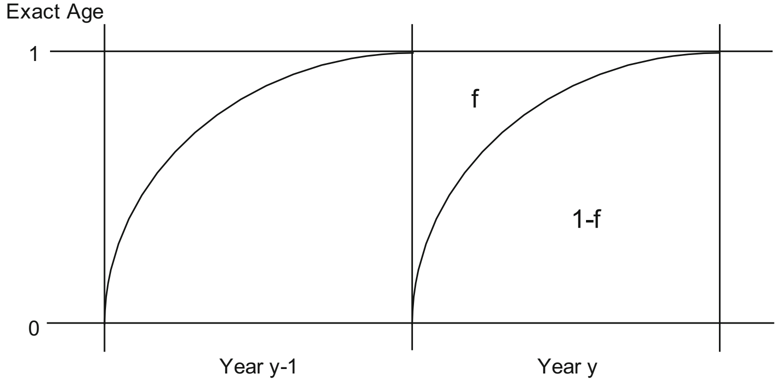


Fig. 4.3 Lexis diagram illustrating how a proportion equal to the separation factor f of deaths at age 0 in a year involve children born the previous year

because it uses both census and vital registration data, in respect of which there may be differential under-enumeration at age 0. If using Eq. 4.9 or Eq. 4.12, both of which require only registration data, to obtain q_0 you should not assume that a value for m_0 obtained from Eq. 4.1 or Eq. 4.2 (under the assumption that $m_0 = M_0$; i.e., that the life table death rate at age 0 equals the observed death rate at age 0) is satisfactory. If you do this m_0 may be affected by inconsistency between the two sources of data while q_0 is not, and the two may be incompatible. ***If either Eq.*** 4.9 ***or Eq.*** 4.12 ***is used to obtain*** q_0***, ensure that your value of*** m_0 ***satisfies Eq.*** 4.11***, under the assumption that*** $m_0 = M_0$. In other words, after rearranging Eq. 4.11 to make M_0 its subject, obtain m_0 from:

$$m_0 = M_0 = q_0 / \left[1 - (1 - f)\, q_0\right] \tag{4.13}$$

Returning to the separation factor f, Keyfitz's formula for f in terms of the death rate at age 0, M_0 (Eq. 4.10), can be used in conjunction with our earlier equation linking q_0 to M_0 (Eq. 4.11) to yield a table of values of f corresponding to values of the infant mortality rate q_0. This is presented as Table 4.1. It can be used to estimate a value for f given the infant mortality rate using linear interpolation (which we will deal with shortly). The introduction of Table 4.1 at this point probably seems rather strange, because we are discussing ways of estimating q_0, some of which use the separation factor f, and yet we are introducing a table that requires us to ***already know*** q_0 in order to obtain f. We would only use Table 4.1 to obtain a value of f for use in generating life table functions ***other than*** q_0. The reason for introducing it here is that it makes an interesting ***theoretical*** point.

The relationship between the infant mortality rate q_0 and the separation factor f in Table 4.1 is ***positive***. As the infant mortality rate increases, so does the separation factor; if q_0 is low f is low, and if q_0 is high f is high. Since f measures the proportion of infant deaths that occur in the year following the year of birth, this relationship

Table 4.1 Values of the Keyfitz separation factor f corresponding to selected values of the infant mortality rate

IMR = q_0	f	IMR = q_0	f
.01	.085	.11	.273
.02	.105	.12	.293
.03	.122	.13	.313
.04	.140	.14	.333
.05	.159	.15	.352
.06	.178	.16	.372
.07	.196	.17	.392
.08	.215	.18	.412
.09	.235	.19	.432
.10	.254	.20	.452

Source: P.F. McDonald, unpublished notes

means that ***the lower the infant mortality rate, the greater is the concentration of infant deaths early in the first year of life, and therefore in the year of birth***. Thus, in a country like Australia in which the infant mortality rate is very low, the few infant deaths that do occur (4.66 per 1,000 live births during 2005–2010) mostly occur soon after birth, and therefore overwhelmingly (well over 90 % according to Table 4.1, since $q_0 = 0.00466 < 0.01$) in the year of birth. But in countries with much higher infant mortality rates (e.g., Chad with 131.17 infant deaths per 1,000 live births during 2005–2010 and Afghanistan with 135.95) the concentration of infant deaths in the year of birth is less marked (more like 65–70 %).

Why is this? A measure of understanding can be achieved by thinking in terms of ***causes*** of death. While specifying cause of death can be complicated in individual cases, with respect to infant deaths a basic distinction can be made between ***endogenous*** and ***exogenous*** causes of death. Endogenous causes of death are those that arise from the genetic makeup of the child and the circumstances of prenatal life and the birth process. Exogenous causes of death reflect the physical, social and medical environment to which a baby is exposed after birth.

Infant deaths due to endogenous causes tend to be heavily concentrated in ***the first few hours and days*** after birth and to be ***not readily preventable***. They are in many ways extensions of ***foetal mortality*** (miscarriage and stillbirth), with conditions like congenital deformities and extreme prematurity often implicated. Those due to exogenous causes are much less concentrated early in the first year of life and are ***more readily preventable***. As infant mortality falls from a high level in a population, it is primarily deaths from exogenous causes occurring later in the first year of life that are eliminated (by basic, often relatively inexpensive, public health initiatives – improved sanitation, educating mothers to avoid, detect and treat child illnesses, providing rudimentary post-natal medical services, etc.). Reducing deaths from endogenous causes, by contrast, tends to depend on medical advances and the availability of sophisticated medical technology, both of which are expensive and not a high priority until deaths that can be prevented more cheaply ***have*** been prevented. Thus, as infant mortality declines endogenous causes of death become

Table 4.2 Values of the separation factor f corresponding to selected values of the infant mortality rate in Coale-Demeny 'West' and United Nations 'General' model life tables

IMR = q_0	Male f	Female f
.006	.061	.070
.008	.067	.076
.01	.072	.081
.02	.100	.110
.03	.128	.138
.04	.156	.169
.05	.185	.198
.06	.214	.229
.07	.243	.260
.08	.272	.290
.09	.303	.321
.10	.329	.350
.11 to .20	.330	.350

Source: Derived from Coale et al. (1983) and United Nations (1982)

proportionately more prominent. Infant deaths accordingly increasingly occur very early in the first year of life, and the separation factor f (the proportion of infant deaths occurring in the year following the year of birth) declines.

An alternative table of separation factors associated with different values of q_0 is presented in Table 4.2. These derive from ***model life tables*** prepared by Coale and Demeny (1966) (see also Coale et al. 1983) and the United Nations (1982) – specifically the Coale-Demeny 'West' and United Nations 'General' model life tables. Although ***based on*** life tables for actual populations, these 'model' life tables relate to ***hypothetical*** populations. They generalize mortality patterns found in ***families*** of actual life tables, their purpose being to aid the analysis of mortality patterns in populations lacking adequate data to construct their own reliable life tables. The Coale-Demeny model life tables are based on 192 of 326 life tables examined (the other 134 were discarded for various reasons) for predominantly Western populations. The 'West' variant (there are also 'North', 'South' and 'East' variants) represents the most general mortality pattern, and reflects populations of Western Europe, North America, Australasia, Japan, Taiwan and White South Africa. The United Nations model life tables, to be distinguished from an earlier set (United Nations 1955), have a developing country focus, their 'General' variant averaging four more regionally specific variants. Despite their respective developed and developing country foci, values of f in relation to values of q_0 underpinning the Coale-Demeny 'West' and UN 'General' model life tables are virtually identical.

Comparing Table 4.2 with Table 4.1 we can note three things. First, again values of f increase as values of q_0 do, but only until $q_0 = 0.10$ (i.e., 100 infant deaths per 1,000 live births). Thereafter they remain constant at higher infant mortality levels. Second, values of f increase more rapidly with increasing q_0 in Table 4.2

to $q_0 = 0.10$, suggesting a more rapid process of concentration of infant deaths in the year of birth as infant mortality declines below 100 deaths per 1,000 live births. Finally, for any value of q_0, f is a little higher for females than for males, meaning that, at any level of q_0, male infant deaths are slightly more concentrated in the year of birth than female infant deaths are. If seeking to obtain a value of f corresponding to a given value of q_0 either Table 4.1 or Table 4.2 could be used. The latter has the attraction of providing sex-specific estimates; the former differentiates at values of $q_0 > 0.10$ which the latter does not. The important thing is to make clear which option has been taken; is f based on the Keyfitz equation or the Coale-Demeny 'West' and UN 'General' model life tables.

Generating a Single-Year-of-Age Life Table: The l_x, $_1d_x$ and $_1p_x$ Columns

Our digression to explain and elaborate on the concept of a separation factor over, we will refocus on the generation of a single-year-of-age life table. Once a set of $_1q_x$ values has been obtained (as in Table 4.3), three other closely related functions can be generated quite simply. Recalling that l_0, the life table radix, is conventionally arbitrarily set at 100,000, we can obtain the l_x, $_1d_x$ and $_1p_x$ columns by applying the following equations, starting with $x = 0$ and increasing its value by 1 for each subsequent application. Note that, for simplicity, leading subscripts have been dropped in these equations.

$$d_x = l_x q_x \tag{4.14}$$

$$l_{x+1} = l_x - d_x \tag{4.15}$$

$$p_x = 1 - q_x \tag{4.16}$$

Other equations may also be useful in manipulating these same life table functions:

$$q_x = d_x / l_x \tag{4.17}$$

$$l_{x+1} = l_x p_x = l_x (1 - q_x) \tag{4.18}$$

$$d_x = l_x - l_{x+1} \tag{4.19}$$

Table 4.3 Single-year-of-age life table for Australian males, 2009–2011

Age (x)	$_1M_x$	$_1q_x$	$_1p_x$	l_x	$_1d_x$	$_1L_x$	T_x	e^o_x
0	0.00498	0.00496	0.99504	100000	496	99555	7974745	79.75
1	0.00041	0.00040	0.99960	99504	41	99481	7875190	79.14
2	0.00019	0.00019	0.99981	99463	19	99453	7775709	78.18
3	0.00014	0.00015	0.99985	99444	14	99437	7676256	77.19
4	0.00013	0.00013	0.99987	99430	13	99423	7576819	76.20
5	0.00012	0.00012	0.99988	99417	12	99411	7477396	75.21
6	0.00010	0.00011	0.99989	99405	10	99400	7377985	74.22
7	0.00010	0.00010	0.99990	99395	10	99390	7278585	73.23
8	0.00010	0.00009	0.99991	99385	10	99380	7179195	72.24
9	0.00009	0.00009	0.99991	99375	9	99371	7079815	71.24
10	0.00009	0.00009	0.99991	99366	9	99362	6980444	70.25
11	0.00009	0.00009	0.99991	99357	9	99353	6881082	69.26
12	0.00010	0.00010	0.99990	99348	10	99343	6781729	68.26
13	0.00012	0.00012	0.99988	99338	12	99333	6682386	67.27
14	0.00016	0.00017	0.99983	99326	16	99319	6583053	66.28
15	0.00025	0.00025	0.99975	99310	25	99298	6483734	65.29
16	0.00036	0.00037	0.99963	99285	36	99268	6384436	64.30
17	0.00047	0.00047	0.99953	99249	47	99226	6285168	63.33
18	0.00054	0.00055	0.99945	99202	54	99176	6185942	62.36
19	0.00061	0.00060	0.99940	99148	60	99118	6086766	61.39
20	0.00063	0.00062	0.99938	99088	62	99057	5987648	60.43
21	0.00063	0.00063	0.99937	99026	62	98995	5888591	59.47
22	0.00063	0.00063	0.99937	98964	62	98933	5789596	58.50
23	0.00064	0.00064	0.99936	98902	63	98871	5690663	57.54
24	0.00065	0.00065	0.99935	98839	64	98807	5591792	56.57
25	0.00067	0.00067	0.99933	98775	66	98742	5492985	55.61
26	0.00070	0.00070	0.99930	98709	69	98675	5394243	54.65
27	0.00074	0.00074	0.99926	98640	73	98604	5295568	53.69
28	0.00077	0.00077	0.99923	98567	76	98529	5196964	52.73
29	0.00081	0.00081	0.99919	98491	80	98451	5098435	51.77
30	0.00085	0.00085	0.99915	98411	84	98369	4999984	50.81
31	0.00090	0.00089	0.99911	98327	88	98283	4901615	49.85
32	0.00094	0.00094	0.99906	98239	92	98193	4803332	48.89
33	0.00099	0.00098	0.99902	98147	97	98099	4705139	47.94
34	0.00103	0.00103	0.99897	98050	101	98000	4607040	46.99
35	0.00108	0.00108	0.99892	97949	106	97896	4509040	46.03
36	0.00114	0.00114	0.99886	97843	111	97788	4411144	45.08
37	0.00120	0.00120	0.99880	97732	117	97674	4313356	44.13
38	0.00126	0.00126	0.99874	97615	123	97554	4215682	43.19
39	0.00132	0.00133	0.99867	97492	129	97428	4118128	42.24
40	0.00141	0.00140	0.99860	97363	137	97295	4020700	41.30
41	0.00149	0.00149	0.99851	97226	145	97154	3923405	40.35
42	0.00159	0.00159	0.99841	97081	154	97005	3826251	39.41

(continued)

Table 4.3 (continued)

Age (x)	${}_1M_x$	${}_1q_x$	${}_1p_x$	l_x	${}_1d_x$	${}_1L_x$	T_x	e^o_x
43	0.00170	0.00170	0.99830	96927	165	96845	3729246	38.47
44	0.00183	0.00183	0.99817	96762	177	96675	3632401	37.54
45	0.00197	0.00197	0.99803	96585	190	96491	3535726	36.61
46	0.00214	0.00213	0.99787	96395	206	96293	3439235	35.68
47	0.00231	0.00231	0.99769	96189	222	96080	3342942	34.75
48	0.00252	0.00252	0.99748	95967	242	95848	3246862	33.83
49	0.00274	0.00274	0.99726	95725	262	95596	3151014	32.92
50	0.00299	0.00299	0.99701	95463	285	95322	3055418	32.01
51	0.00326	0.00325	0.99675	95178	310	95025	2960096	31.10
52	0.00355	0.00354	0.99646	94868	336	94702	2865071	30.20
53	0.00385	0.00384	0.99616	94532	363	94353	2770369	29.31
54	0.00417	0.00416	0.99584	94169	392	93976	2676016	28.42
55	0.00452	0.00451	0.99549	93777	423	93568	2582040	27.53
56	0.00490	0.00488	0.99512	93354	456	93129	2488472	26.66
57	0.00530	0.00529	0.99471	92898	491	92656	2395343	25.78
58	0.00574	0.00573	0.99427	92407	529	92146	2302687	24.92
59	0.00626	0.00623	0.99377	91878	573	91595	2210541	24.06
60	0.00682	0.00680	0.99320	91305	621	90999	2118946	23.21
61	0.00746	0.00744	0.99256	90684	674	90351	2027947	22.36
62	0.00819	0.00815	0.99185	90010	734	89648	1937596	21.53
63	0.00901	0.00897	0.99103	89276	801	88881	1847948	20.70
64	0.00993	0.00988	0.99012	88475	874	88045	1759067	19.88
65	0.01096	0.01090	0.98910	87601	955	87131	1671022	19.08
66	0.01211	0.01204	0.98796	86646	1043	86132	1583891	18.28
67	0.01341	0.01332	0.98668	85603	1140	85041	1497759	17.50
68	0.01484	0.01473	0.98527	84463	1244	83850	1412718	16.73
69	0.01643	0.01629	0.98371	83219	1356	82551	1328868	15.97
70	0.01815	0.01800	0.98200	81863	1473	81137	1246317	15.22
71	0.02009	0.01989	0.98011	80390	1599	79601	1165180	14.49
72	0.02220	0.02195	0.97805	78791	1730	77937	1085579	13.78
73	0.02456	0.02426	0.97574	77061	1870	76138	1007642	13.08
74	0.02725	0.02689	0.97311	75191	2022	74194	931504	12.39
75	0.03034	0.02990	0.97010	73169	2187	72090	857310	11.72
76	0.03392	0.03335	0.96665	70982	2368	69814	785220	11.06
77	0.03803	0.03733	0.96267	68614	2561	67350	715406	10.43
78	0.04276	0.04188	0.95812	66053	2766	64688	648056	9.81
79	0.04819	0.04707	0.95293	63287	2979	61816	583368	9.22
80	0.05437	0.05296	0.94704	60308	3193	58729	521552	8.65
81	0.06143	0.05960	0.94040	57115	3405	55430	462823	8.10
82	0.06935	0.06705	0.93295	53710	3601	51925	407393	7.59
83	0.07828	0.07536	0.92464	50109	3776	48235	355468	7.09
84	0.08828	0.08456	0.91544	46333	3918	44384	307233	6.63
85	0.09935	0.09467	0.90533	42415	4015	40414	262849	6.20

(continued)

Table 4.3 (continued)

Age (x)	$_1M_x$	$_1q_x$	$_1p_x$	l_x	$_1d_x$	$_1L_x$	T_x	$e^o{}_x$
86	0.11145	0.10557	0.89443	38400	4054	36374	222435	5.79
87	0.12444	0.11713	0.88287	34346	4023	32329	186061	5.42
88	0.13815	0.12920	0.87080	30323	3917	28353	153732	5.07
89	0.15258	0.14166	0.85834	26406	3741	24518	125379	4.75
90	0.16798	0.15479	0.84521	22665	3509	20890	100861	4.45
91	0.18491	0.16908	0.83092	19156	3238	17511	79971	4.17
92	0.19999	0.18148	0.81852	15918	2889	14446	62460	3.92
93	0.22132	0.19888	0.80112	13029	2591	11707	48014	3.69
94	0.24273	0.21582	0.78418	10438	2253	9282	36307	3.48
95	0.26433	0.23257	0.76743	8185	1904	7203	27025	3.30
96	0.28050	0.24484	0.75516	6281	1538	5483	19822	3.16
97	0.29555	0.25623	0.74377	4743	1215	4111	14339	3.02
98	0.31104	0.26755	0.73245	3528	944	3035	10228	2.90
99	0.32563	0.27815	0.72185	2584	719	2208	7193	2.78
100+	0.37412	1.00000	0.00000	1865	1865	4985	4985	2.67

Generating a Single-Year-of-Age Life Table: The $_1L_x$, T_x and $e^o{}_x$ Columns

We noted earlier that the $_1L_x$ (often abbreviated to L_x) column of the life table gives the number of person-years lived between exact ages x and x + 1 by members of the life table radix population who survive to exact age x. Another way of putting this is that it gives the number of person-years of exposure to the risk of dying between exact ages x and x + 1 for the life table population. This statement should sound familiar; it should sound to you like the specification of the ***denominator*** of some demographic ***rate***, and it is. We can note in passing that the life table death rate $_1m_x$ (often abbreviated to m_x) is given by:

$$m_x = d_x / L_x \tag{4.20}$$

Another interpretation of the L_x column is that it gives ***the age structure of the life table stationary population***. We mentioned this concept of the 'life table population' and noted that it was a stationary population in passing earlier in this chapter. The L_x column gives the life table population. A population experiencing 100,000 births annually (the life table radix), which was subject to the probabilities of dying at each age specified in the q_x column (which would result in exactly 100,000 deaths annually, since these q_x values cause the life table radix population to completely die out), and which was closed to migration would have constant numbers in each single-year age group equal to the schedule of L_x values.

The general equation for calculating L_x is:

$$L_x = f_x l_x + (1 - f_x)\, l_{x+1} \tag{4.21}$$

Where f_x = a separation factor appropriate to the age group bounded by exact ages x and x + 1.

However, when $x \neq 0$ (i.e., for all values of x ***except*** $x = 0$), the value of f_x normally is ½, because it is reasonable to assume an even distribution of deaths by exact age, otherwise known as ***linear survivorship***, between exact ages x and x + 1. Thus, except where $x = 0$, Eq. 4.21 reduces to:

$$L_x = \tfrac{1}{2}\,(l_x + l_{x+1}) \tag{4.22}$$

It does sometimes happen with published life tables that you are unable to reproduce given values of L_x from given values of l_x using Eq. 4.22. You are most likely to encounter this situation at early childhood ages in life tables for high mortality populations, when an argument similar to that on which the use of a separation factor of less than ½ at age 0 is based may be advanced for immediately older ages as well. Non-use of Eq. 4.22 at these ages is an indication that, within the relevant single-year age interval, deaths concentrate at its lower end, and a separation factor with a value less than ½ has been used (e.g., the values 0.43, 0.45, 0.47 and 0.49 which Chiang (1984) uses for x = 1, 2, 3 and 4 respectively).

We should also pause at this point over the introduction of the concept of 'linear survivorship'. We noted earlier that separation factors are used at ages other than 0, especially in the construction of abridged life tables, and the notion of linear survivorship provides a useful way of specifying why, in general, separation is an issue in life table construction. ***Separation factors are used in life table construction to adjust for non-linear survivorship over an age interval***. With single-year-of-age life tables this problem occurs mainly at age 0, although as just indicated, it may also occur at slightly older childhood ages. As we shall see, though, it is a more general issue in the construction of abridged life tables because of the wider age intervals over which survivorship is measured.

Resuming our discussion of L_x values, for x = 0 we have:

$$L_0 = f l_0 + (1 - f)\, l_1 \tag{4.23}$$

Where f, as previously, is the separation factor appropriate to mortality during the first year following birth.

If data are available that classify infant deaths by both year of occurrence and year of birth it is possible to calculate f directly (and absolutely accurately) as the proportion of infant deaths of children born in year y that occurred in year y + 1. However, it is usual to have data classified only by year of occurrence, whence we must estimate f. In using Eq. 4.23 we might wish to obtain a value for f from either of Tables 4.1 and 4.2. These tables specify f values for only a restricted number of values of the infant mortality rate (or q_0), but as indicated earlier, ***linear interpolation*** can be used to estimate f values corresponding to other q_0 values. Linear interpolation assumes that relationships between f and q_0 between successive pairs of q_0 and f values in Tables 4.1 and 4.2 are linear (i.e., follow a straight line). To apply this principle, identify in either table the values of q_0 nearest to, but lower than, and nearest to, but higher than, your observed q_0. These define the IMR interval within which interpolation will occur. Then evaluate the equation:

$$f = f_l + \{[(IMR_o - IMR_l) / (IMR_u - IMR_l)] . (f_u - f_l)\} \tag{4.24}$$

Where IMR_o = the ***observed*** value of q_0; IMR_l and IMR_u = the values of q_0 defining the ***lower*** and ***upper*** limits, respectively, of the IMR interval in Table 4.1 or 4.2 within which IMR_o lies; f_l and f_u are the values of f associated with IMR_l and IMR_u, respectively, in Table 4.1 or 4.2.

An alternative approach to calculating L_x values uses the equation:

$$L_x = d_x/M_x \tag{4.25}$$

This formula can be used for all ages, including age 0 provided that q_0 has been obtained from M_0. Its use assumes equality between observed death rates M_x and life table death rates m_x. A rearrangement of Eq. 4.25 to make M_x its subject and a comparison of the result with Eq. 4.20 should convince you of this.

We turn next to the function T_x, which gives total person-years of exposure of the life table population to the risk of dying at all ages beyond exact age x. It is of little use in itself, but vital to the calculation of average expectations of life remaining ($e^o{}_x$ values), measures which are very useful in comparing mortality levels in different populations. It is also valuable in simplifying problems which require calculation of the number of person-years of exposure to the risk of dying (or the number of persons in the life table stationary population) between exact ages x and $x + n$ (where $n > 1$). Rather than having to sum L_x values over the range x to $x + n - 1$ (a tedious process if n is not small) we need only subtract T_{x+n} from T_x. That is:

$$\Sigma_{i=x,x+n-1} L_i = T_x - T_{x+n}$$

Where n gives the width in years of the age interval over which person-years of exposure to risk is required.

The basic equation for T_x is:

$$T_x = \Sigma_{i=x,\omega} L_i$$

Where ω = the oldest age reached by anyone in the life table population.

Normal practice in generating T_x values, however, is to begin at the ***bottom*** of the life table (i.e., at the ***oldest*** age reached by the life table population, where $T_x = L_x$)

and work from there to the top of the table (age 0) shuffling back and forth between the T_x and L_x columns using the equation:

$$T_x = T_{x+1} + L_x \tag{4.26}$$

And so to the $e^o{}_x$ column. The concept of average expectation of life ***at birth*** ($e^o{}_0$) probably has a reasonable level of public recognition in developed populations, but the single-year-of-age life table also provides measures of average expectation of ***life remaining*** at each subsequent birthday as well. These address the question, 'Given that a man or woman has survived to exact age (birthday) x, how many more years, on average, can he or she expect to live?' The calculation is very simple; total person-years lived by the life table population beyond exact age x (T_x) is apportioned among survivors at exact age x (l_x). Hence:

$$e^o{}_x = T_x / l_x \tag{4.27}$$

Always remember, though, that $e^o{}_x$ is ***an average***; some persons attaining age x live longer, others not as long, and very few live for exactly $e^o{}_x$ further years. Values from this column are very useful in comparing the general health of populations (as this is manifested by their survival capacities) and in monitoring changes in general health levels over time. They are measures of ***longevity*** (i.e., length of life).

Generating a Single-Year-of-Age Life Table: Other Issues

In this subsection we deal with a number of miscellaneous issues that arise in relation to the generation of single-year-of-age life tables. We focus on the bottom of the life table (the oldest ages); taking migration into account in the calculation of q_x values; and the calculation of the separation factor f from detailed data on infant deaths by age.

The Bottom of the Life Table

Particular difficulties can arise in generating single-year-of-age life tables at the oldest ages. Because so few people live to very advanced ages (ages greater than 100 years, say), we can strike phenomena such as: (i) single-year age groups in which there was a death (or more) during the year in question, but nobody at risk of dying at mid-year, so that a value for M_x, and hence one for q_x, cannot be calculated; (ii) single-year age groups which yield q_x values equal to or greater than 1 (implying that all survivors to exact age x die at age x) when there are also deaths and persons at risk at older ages; and (iii) single-year age groups in which there are neither deaths nor persons at risk, when there ***are*** deaths and persons at risk at older ages.

Sometimes this sort of problem is dealt with by truncating the life table at some single-year age group beyond which difficulties of the type described occur. In other words, entries for the final few single-year age groups are simply omitted. A second strategy is to group the final few ages, creating an age group x+, or 'x and over'. Under this strategy, for the last line of the life table $q_x = 1.0$, since everyone attaining exact age x dies at some subsequent age. Also:

$$L_{x+} = T_x = l_x/m_{x+} = l_x/M_{x+} \tag{4.28}$$

Where M_{x+} is the observed death rate for the age group stretching from x to whatever the final age featuring in our data happens to be, and is assumed to equal m_{x+}.

Other life table functions can then be obtained by conventional means.

A common approach of this type to dealing with the bottom of a life table is to group all ages above exact age 85. This approach is frequently used in the construction of abridged life tables in particular, but may also be used to complete single-year-of-age tables, especially for high mortality populations in which relatively few people survive into their late 80s and beyond. Clearly $q_{85+} = 1.0$, and ***provided*** $\boldsymbol{l_0 = 100{,}000}$ we have as an approximation:

$$L_{85+} = T_{85} \cong l_{85}.\log_{10} l_{85} \tag{4.29}$$

The approximation this equation provides for T_{85} (note that $\log_{10} l_{85}$ means 'logarithm of l_{85} to base 10') is not, however, always a good one. If it is used to estimate T_{85} in Table 4.3, for example, it yields an answer of 186,168 compared to the value shown in the table of 262,849. Table 4.3 is a life table for a low mortality population (expectation of life at birth 79.8 years), not the type of population for which grouping above exact age 85 would ordinarily be undertaken. Better approximations seem to be provided by Eq. 4.29 for higher mortality populations of developing countries. Considering 1970, 1980 and 1990 female life tables for the Philippines with life expectancies at birth of respectively 61.5, 65.1 and 67.4 years constructed by Flieger and Cabigon (1994), for example, values of T_{85} (hence L_{85+}) of 61,711, 70,439 and 95,979 compare with estimates from Eq. 4.29 of 62,286, 70,473 and 81,480 respectively. Equation 4.29 relies on an observed empirical 'regularity' rather than a theoretical relationship. Other equations can be used, including Eq. 4.28 with x set equal to 85. If all of this sounds disquietingly imprecise, there is some consolation in the fact that the degree of inaccuracy that might exist in life table functions at these advanced ages has little impact on functions at younger ages, even though there is a feedback to those ages through the T_x and $e^o{}_x$ columns.

Taking Migration into Account

Equation 4.6 above indicated that values of q_x could be obtained from:

$$q_x = D_x / \left(P_x + {}^1\!/_2 D_x\right)$$

Where D_x = deaths at age x during the year for which q_x is required; P_x is the mid-year population aged x in that year.

This equation, however, assumes a closed population. It is often argued that this is usually not an unreasonable assumption, and that adjustment for migration is rarely called for. But where migration, and more especially ***net*** migration, is significant, an adjustment for it probably should be made in calculating q_x values. Denoting out-migrants aged x during the year for which a life table is being constructed by ${}_1o_x = o_x$, and in-migrants by ${}_1i_x = i_x$, the necessary adjustment is achieved by using in place of Eq. 4.6 the following alternative:

$$q_x = D_x / \left[P_x + {}^1\!/_2 (D_x + o_x - i_x)\right] \tag{4.30}$$

Calculating Separation Factors from Detailed Data on Infant Deaths by Age

Earlier discussion of the separation factor f claimed that it summarized the distribution of infant deaths between the year of birth and the following year, and relied on approximations for f obtained from the Keyfitz equation or model life tables which linked f to q_0 (Tables 4.1 and 4.2). As illustrated in Fig. 4.3 an alternative, essentially equivalent but strictly more correct conceptualization says that ***f apportions infant deaths in a year between those involving children born in that year (1 − f) and in the PREVIOUS year (f)***. Moreover, where detailed data on infant deaths by age such as are presented in Table 4.4 are available, it is possible to use Lexis diagram principles to calculate the separation factor f empirically, thus obtaining a value likely to be more accurate than other estimation techniques will yield.

'Hang on', you say. 'Didn't we decide that Lexis diagram principles were invalid at age 0?' Indeed we did, but that was in the context of treating the first year of life as a single age interval. Deaths were so heavily concentrated early in the interval that to assume they were evenly distributed through it by exact age was unrealistic. Having data such as those in Table 4.4, however, allows us to ***slice the first year of life into a series of much narrower age intervals*** – seven that are each only 1 day wide (covering the first week after birth), three more that are each 1 week wide (covering the rest of the first month after birth), and eleven that are each 1 month wide (covering the remainder of the first year after birth). While Lexis diagram assumptions are unreasonable when applied over the whole of the first year following birth, they are much more reasonable when applied ***separately within each of the 'slices' just defined***. That is, it is reasonable to assume that deaths aged less than 1 day were evenly spread through the first 24 h after birth, those aged 1 day were evenly spread through the second 24 h, . . ., those aged 1 week were evenly

Table 4.4 Calculation of the separation factor f for Australian infant deaths during 2010 using detailed data on deaths by age

			Persons dying	
Age at death	Mean age at death (years)	Deaths in 2010	Born 2009	Born 2010
<1 day	${}^{1}/_{730} = 0.001370$	382	1	381
1 day	${}^{3}/_{730} = 0.004110$	68	0	68
2 days	${}^{5}/_{730} = 0.006849$	47	0	47
3 days	${}^{7}/_{730} = 0.009589$	33	0	33
4 days	${}^{9}/_{730} = 0.012329$	25	0	25
5 days	${}^{11}/_{730} = 0.015068$	18	0	18
6 days	${}^{13}/_{730} = 0.017808$	8	0	8
7–13 days	${}^{21}/_{730} = 0.028767$	61	2	59
14–20 days	${}^{35}/_{730} = 0.047945$	35	2	33
21–27 days	${}^{49}/_{730} = 0.067123$	30	2	28
1 month	${}^{3}/_{24} = 0.125000$	68	9	59
2 months	${}^{5}/_{24} = 0.208333$	59	12	47
3 months	${}^{7}/_{24} = 0.291667$	41	12	29
4 months	${}^{9}/_{24} = 0.375000$	28	11	17
5 months	${}^{11}/_{24} = 0.458333$	24	11	13
6 months	${}^{13}/_{24} = 0.541667$	17	9	8
7 months	${}^{15}/_{24} = 0.625000$	14	9	5
8 months	${}^{17}/_{24} = 0.708333$	11	9	2
9 months	${}^{19}/_{24} = 0.791667$	10	8	2
10 months	${}^{21}/_{24} = 0.875000$	9	8	1
11 months	${}^{23}/_{24} = 0.958333$	10	10	0
Total		998	115	883
Separation factor = 115/998 = 0.115				

spread through days 8–14 following birth, . . . , those aged 1 month were evenly spread through the second month following birth, etc.

Using these assumptions it is possible, with reasonable precision, to allocate infant deaths in each age 'slice' between those of children born during the year in which death occurred and those of children born the previous year. This is the type of exercise performed in Table 4.4, and its basis is illustrated in Fig. 4.4. This Lexis diagram shows ***a portion*** of one square of the conventional Lexis grid; a portion covering (i) the months of January and February in the year (2010) for which we wish to calculate the separation factor f, and (ii) demographic experience between exact ages 0 and 2 months. Try to imagine ***the remainder*** of this grid square – another ten monthly divisions along each axis (March, April, . . . , December along the horizontal axis and exact ages 3 months, 4 months, . . . , 12 months up the vertical axis).

The horizontal lines through Fig. 4.4 divide mortality at age 0 during 2010 into the 'slices' just described. The bottom 'slice' corresponds to the top row of Table 4.4 (deaths aged <1 day), the next 'slice' corresponds to the next row (deaths aged

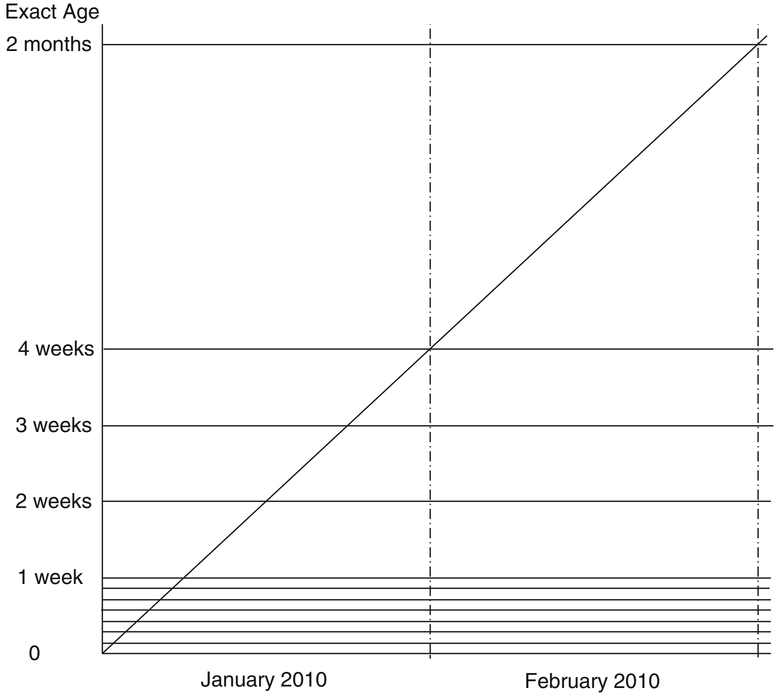


Fig. 4.4 Lexis diagram illustrating calculation of the separation factor f for Australian infant deaths during 2010 using detailed data on deaths by age

1 day), and so on. The diagonal line through the diagram is the life line that separates the demographic (in this case mortality) experience of the 2010 birth cohort (to the right of the line) from that of the 2009 birth cohort (to the left of it).

Consider the bottom 'slice' in Fig. 4.4; that representing deaths aged <1 day. Most of it lies to the right of the diagonal dividing the 2009 birth cohort from the 2010 birth cohort; i.e., most deaths aged <1 day during 2010 involved babies also born in 2010. All that lies ***to the left*** of the diagonal is a small triangle. This triangle represents the deaths aged <1 day during 2010 that involved babies born during 2009. What fraction of the area of the entire 'slice' is the area of this triangle? Well, the slice is 1 day wide on the vertical scale, and the diagonal forming the hypotenuse of the triangle runs at 45°, implying that the triangle is half of a square that also extends 1 day along the horizontal axis. The 'slice' comprises 365 such squares – one for each day of 2010. Half of $^1/_{365}$ of the 'slice' is $^1/_{730}$ of the 'slice', and this fraction is the first entry in the 'Mean age at death' column of Table 4.4. We estimate that $^1/_{730}$ of the deaths aged <1 day during 2010 involved children born during 2009. As there were 382 such deaths, calculating to the nearest whole death (since dealing with fractions of deaths makes no intuitive sense) we conclude that 1 of them occurred to a member of the 2009 birth cohort and the other 381 to members of the 2010 birth cohort (top row of Table 4.4).

Moving to the second 'slice' – that corresponding to deaths aged 1 day – the area to the left of the diagonal consists of a complete '1 day by 1 day' square plus a triangle half the area of another such square. That is, $^{1}/_{365} + ^{1}/_{730} = ^{3}/_{730}$ of the deaths aged 1 day during 2010 were of children born during 2009. With only 68 deaths in total to be apportioned between the two birth cohorts, ***to the nearest whole death NONE*** of them involved a child born during 2009; all involved children born during 2010 (second row of Table 4.4).

And so on. When we get to the wider 'slice' corresponding to deaths aged 1 week (but less than 2 weeks) the entire slice consists of 365 rectangles 1 day wide by 1 week high placed side by side, and the area to the left of the diagonal consists of the first 7 of these plus half of the next 7; i.e., $^{7}/_{365} + ^{1}/_{2}(^{7}/_{365}) = ^{14}/_{730} + ^{7}/_{730} = ^{21}/_{730}$ of the deaths aged 1 week, or 7–13 days, were deaths of children born during 2009 (row 8 of Table 4.3). When we have completed apportioning deaths in each age 'slice' during 2010 between the 2009 and 2010 birth cohorts in this fashion, we add the deaths allocated to the earlier (2009) cohort and express the answer as a proportion of total infant deaths during 2010.

This is our separation factor. Its value (Table 4.4) is somewhat different from that we would have calculated using the Keyfitz equation (Eq. 4.10) (0.115 compared to 0.076). This emphasizes that estimates provided by approaches such as Keyfitz's are just that – estimates, and not always very good ones. Fortunately, some imprecision in the value of the separation factor f usually is of limited consequence for the value of q_0 (Shryock, Siegel, and Associates 1973), and hence for the values of other life table functions. This is just as well, because it is relatively rare to have data as detailed as those in Table 4.4, whence we are ***forced*** to use approximations for f such as those presented in Tables 4.1 and 4.2 (the literature contains a range of approaches to calculating separation factors, and to fully understand how any particular life table was constructed it is necessary to check how separation was done). Imprecision in f does, for example, tend to be of less consequence for q_0 than using Eq. 4.11 rather than Eq. 4.12 as the basic calculating equation; i.e., it tends to be less of a problem than the former equation's reliance on a mixture of registration and census data, between which there may be different levels of under-enumeration at age 0.

One final point about Table 4.4 that may require some clarification is the heading 'Mean age at death' used for the column in which the fractions of deaths in each age 'slice' to be allocated to births during the previous year are given. It happens that these fractions equal the exact ages, in fractions of a year, that form the mid-points of the age intervals defining the various age 'slices'. And since we are assuming even distributions of deaths by exact age within each 'slice', these mid-points correspond to mean ages at death for those dying in each age interval. For example, children dying aged <1 day (the bottom 'slice' in Fig. 4.4) are assumed to live on average for half a day, which is $^{1}/_{730}$ of a year (top row of Table 4.4); those dying aged 1 day are assumed to live for one and a half days on average, or $^{3}/_{730}$ of a year (second row of Table 4.4), etc. Thus we have an alternative interpretation of these fractions to the one based on areas on the Lexis diagram. ***As a general rule, under***

the assumptions of the Lexis diagram, the mean age in years of infants dying in any age interval gives the proportion of deaths in that interval during a year which involved children born during the previous year. Of course, the qualifier 'under the assumptions of the Lexis diagram' in this statement is all-important. The value of the principle just enunciated depends on age intervals being used which are not so broad as to render these assumptions unreasonable. For example, it is clear from Table 4.4 that there is nothing like an even distribution of infant deaths by exact age at death through the first week following birth, so that to treat the first week as a single age interval and apply the 'general rule' above would be to invoke an assumption that plainly is invalid, and therefore to potentially introduce significant error into one's separation factor. Note also the word 'infants' in the italicized statement above. It says that we are talking ***only*** about deaths between exact ages 0 and 1.

Recipe for Constructing a Single-Year-of-Age Life Table

Having now completed our discussion of the single-year-of-age life table we can set out the steps to be followed in constructing such a life table in recipe form. Assuming, for a given sex in a given year, that we have single-year-of-age data on deaths (${}_1D_x$) and population at mid-year (${}_1P_x$), and perhaps data on live births (B) during the year in question and the previous year, we can proceed as follows:

1. Calculate values of M_x using either Eq. 4.1 or Eq. 4.2.
2. Calculate the separation factor for age 0, f, either directly from a detailed distribution of infant deaths by age in the manner illustrated by Table 4.4 or, if such data are not available, by using an empirical formula such as Keyfitz's equation (Eq. 4.10).
3. Calculate q_0 either using Eq. 4.9 or Eq. 4.12, in which case Eq. 4.13 should be used to obtain m_0, or using Eq. 4.11, in which case $m_0 = M_0$.
4. Calculate values of q_x for $x > 0$ using Eq. 4.7, and set $m_x = M_x$ for each value of x.
5. Set $l_0 = 100{,}000$ and calculate values of l_x for $x > 0$ using Eq. 4.18.
6. Use Eq. 4.19 to calculate values of d_x.
7. Calculate values of L_x using Eq. 4.23 for $x = 0$ and Eq. 4.22 for $x > 0$. Alternatively Eq. 4.20 may be rearranged to give $L_x = d_x / m_x$ for all values of x.
8. Use Eq. 4.28 to calculate L_{x+} and T_x for the last row of the life table.
9. Derive the remainder of the T_x-column using Eq. 4.26, working from the last row of the life table upwards.
10. Calculate values of $e^o{}_x$ using Eq. 4.27.

Abridged Life Tables

It is not always possible, or desirable, to construct single-year-of-age life tables. Available data may not be of sufficient quality to warrant, or indeed even to permit, their construction. Single-year-of-age data may simply not be available. Or they may be so contaminated by age misstatement and/or other defects as to be unusable at a single-year-of-age level of detail. These practical problems create one incentive for constructing abridged life tables. A second important incentive is that single-year-of-age tables are far too detailed for most demographic purposes. Other than at age 0 demographers rarely make comparisons of mortality, whether across populations or through time, by single years of age. Similarly, they frequently use abridged life tables in population projection exercises.

As was indicated earlier, abridged life tables summarize a population's prospects of survival over ***fewer, longer age intervals*** – typically 5-year intervals, with the exception that age 0 is split from ages 1–4 and the oldest ages are grouped into a broad open interval, frequently 85 years and over (85+). Other schemes of abridgement are possible (for example, treating ages 0, 1 and 2–4 separately, 10-year intervals between exact ages 5 and 85, or specifying the terminal open interval differently), but the one just described is the most common. An abridged life table of this type is almost as accurate as a complete life table, provides all the information needed for most demographic analysis purposes, and requires less data and, if the job is being done manually, less 'donkey' work to construct.

In introducing the functions (columns) of a single-year-of-age life table a distinction was drawn between those that pertained to ***exact age*** x (l_x, e^o_x and μ_x) or ***all ages beyond*** exact age x (T_x), and therefore didn't take a prefixed subscript of 1, and those that pertained to the 1-year ***age interval*** x (the interval from exact age x to exact age x + 1), which strictly speaking did take a prefixed subscript of 1 (${}_1q_x$, ${}_1d_x$, ${}_1L_x$, ${}_1m_x$ and ${}_1p_x$). The former functions are sometimes referred to as the ***stock*** functions of the life table, and the latter as the ***flow*** functions. In theory stock functions in an abridged life table remain the same as in the equivalent single-year-of-age life table, and can be directly transferred from the latter to the former. All that happens is that only values for selected x's are transferred. However, if you compare an abridged life table with an equivalent single-year-of-age life table you will often find departures from this theoretical equality. If you do, the major reason is likely to be the manner in which separation has been undertaken. As has already been intimated, separation becomes an issue at all ages in the construction of abridged life tables, and the 'departures' referred to typically arise from errors introduced by the use of a separation factor of one-half beyond exact age 5 in generating abridged life tables. This is a frequently adopted approach which assumes linear survivorship over 5-year age intervals, but its straightforwardness does sacrifice some precision.

Flow functions in an abridged life table (${}_nq_x$, ${}_nd_x$, ${}_nL_x$, ${}_nm_x$ and ${}_np_x$) have no counterpart in a corresponding complete life table, except where n = 1 (generally the case only in the first row of an abridged life table). Other than when n = 1 they cannot be transferred directly from a complete to an abridged life table. The notation

Table 4.5 Abridged life table for Australian males, 2009–2011

Age (x)	n	$_nM_x$	$_nq_x$	l_x	$_nd_x$	$_nL_x$	T_x	$e^o{}_x$
0	1	0.00498	0.00496	100000	496	99561	8014468	80.14
1	4	0.00021	0.00082	99504	82	397722	7914907	79.54
5	5	0.00012	0.00058	99422	58	496965	7517185	75.61
10	5	0.00011	0.00056	99364	56	496680	7020220	70.65
15	5	0.00046	0.00230	99308	229	495968	6523540	65.69
20	5	0.00062	0.00309	99079	306	494630	6027572	60.84
25	5	0.00074	0.00367	98773	362	492960	5532942	56.02
30	5	0.00094	0.00469	98411	461	490903	5039982	51.21
35	5	0.00118	0.00587	97950	575	488313	4549079	46.44
40	5	0.00158	0.00787	97375	767	484958	4060766	41.70
45	5	0.00229	0.01140	96608	1101	480288	3575808	37.01
50	5	0.00344	0.01709	95507	1632	473455	3095520	32.41
55	5	0.00520	0.02568	93875	2411	463348	2622065	27.93
60	5	0.00800	0.03921	91464	3586	448355	2158717	23.60
65	5	0.01303	0.06313	87878	5548	425520	1710362	19.46
70	5	0.02179	0.10331	82330	8506	390385	1284842	15.61
75	5	0.03776	0.17250	73824	12735	337283	894457	12.12
80	5	0.06758	0.28907	61089	17659	261298	557174	9.12
85	ω	0.14678	1.00000	43430	43430	295876	295876	6.81

used for these flow functions has been explained before, but to reiterate, ‘x’ gives the exact age marking the lower limit of the age interval to which the value of a function pertains, and ‘n’ gives the width of the interval in years (so that its upper limit is exact age $x + n$). Table 4.5 gives values of x and n pertaining to each row of the life table in its first two columns (often only x values are given and the user is left to deduce the n values). The final n value shown is ‘ω‘, which means an interval width such that, added to x, it defines the youngest exact age in completed years beyond which no-one lives.

One interesting difference concerning flow functions in an abridged life table is that values in the $_nm_x$ (or $_nM_x$) and $_nq_x$ columns are substantially different (Table 4.5), whereas in the single-year-of-age life table values of $_1m_x$ and $_1q_x$ (m_x and q_x) were similar and often identical (Table 4.3). Why the change? Values of $_1m_x$ and $_nm_x$ are annual life table death rates, respectively for age groups bounded by exact ages x and $x + 1$ and by exact ages x and $x + n$. Values of $_1q_x$ and $_nq_x$ are probabilities of dying between exact ages x and $x + 1$, and exact ages x and $x + n$, respectively. In the case of $_1m_x$ and $_1q_x$ we are comparing an ***annual*** death rate for a single-year age group with a probability of dying over the equivalent ***1-year*** phase of the life cycle, so near-equality is not surprising. But in the case of $_nm_x$ and $_nq_x$ we are comparing an ***annual*** death rate (albeit for an n-year age group) with a probability of dying ***over an n-year period***. Any individual takes n years to live through the relevant age group, and is exposed afresh to the death rate $_nm_x$ ***in each one*** of those years, but $_nq_x$ is a ***single*** measure of a person’s prospects of dying ***over***

the entire n-year phase of the life cycle. Thus, ${}_nq_x$ is expected to be of the order of n-times larger than ${}_nm_x$; in effect it is equivalent to experiencing the death rate ${}_nm_x$ n times – once for each year it takes to age from exact age x to exact age $x + n$.

Essentially, then, values of ${}_nq_x$ in an abridged life table are larger than values of ${}_1q_x$ in a single-year-of-age life table because they give probabilities of dying over longer periods of the life cycle. Similarly values of ${}_nd_x$ and ${}_nL_x$ are larger than values of ${}_1d_x$ and ${}_1L_x$ (giving life table deaths and person-years of exposure to the risk of dying, respectively, over longer periods of the life cycle). Values of ${}_np_x$ are, however, smaller than values of ${}_1p_x$ because the probability of surviving decreases as the length of the age interval for which it is calculated increases.

It was noted above that abridged life tables usually terminate with a broad open age interval. This terminal open interval may begin at exact age 75, 80, 85, 90 or 95, and on occasion even at a younger age than 75. Methods for dealing with the bottom of a life table have already been discussed, and grouping the final ages was noted as being one option. When choosing this option the main challenge was to obtain a value for L_{x+}. We can do this using Eq. 4.28 or, if grouping beyond exact age 85, Eq. 4.29. The latter can ***only*** be used when $x = 85$ (i.e., you ***cannot***, for example, substitute '80' for '85' and '80+' for '85+' in Eq. 4.29), and Eq. 4.28 is in any case preferable to Eq. 4.29. Note that, given that for a terminal open age group $l_x = d_{x+}$, Eq. 4.28 can be extended as follows:

$$L_{x+} = l_x/M_{x+} = d_{x+}/M_{x+}$$

If you compare Table 4.5, the abridged life table for Australian males for 2009–2011, with Table 4.3, the equivalent single-year-of-age life table, you will note that the average expectations of life at birth ($e^o{}_0$) are slightly different (80.14 years compared to 79.75 years). This is a difference of 0.39 years. Now compare values of $e^o{}_{85}$ (the expectation of life remaining for survivors at exact age 85). They are respectively 6.81 and 6.20 years, a difference in the same direction of 0.61 years. This suggests that the higher expectation of life at birth in the abridged life table is largely a product of inaccuracy in the estimate of $L_{85+} = T_{85}$ compared to what single-year-of-age data yield. This reverberates up the life table as ${}_nL_x$ values are added for successively younger ages to generate the T_x column used in calculating $e^o{}_x$ values, with other bits and pieces of error arising from the assumption of linear survivorship over 5-year age intervals partly offsetting this inaccuracy along the way. The terminal age group in Table 4.5 has been set at 85+ because that is the most common terminal age group to encounter in abridged life tables. However, because Australian males in 2009–2011 had low mortality, and therefore a high proportion (over 43 %) surviving to exact age 85, it would have made sense to have used, say, 95+ as the terminal age group. Ideally the terminal age group in an abridged life table should commence at an exact age to which only a small proportion of the population survives, and Table 4.3 shows only a little over 8 % of Australian males in 2009–2011 surviving to exact age 95.

Calculating Abridged Life Tables from Single-Year-of-Age Life Tables

Should the need and the opportunity arise, we can readily collapse a complete life table into an abridged life table. We may wish to do this either to present the essential features of a population's mortality regime more succinctly or to facilitate comparison with some second population for which only an abridged life table is available.

Functions l_x, T_x and $e^o{}_x$ do not change. It is simply a matter of extracting relevant values from the single-year-of-age life table once the age intervals to be used in the abridged life table have been selected. Other functions can then be obtained from the following equations:

$$_nq_x = 1 - (l_{x+n}/l_x) \tag{4.31}$$

$$_np_x = 1 - {}_nq_x = l_{x+n}/l_x \tag{4.32}$$

$$_nd_x = d_x + d_{x+1} + \cdots + d_{x+n-1} = l_x - l_{x+n} = l_x \cdot {}_nq_x \tag{4.33}$$

$$_nL_x = L_x + L_{x+1} + \cdots + L_{x+n-1} \tag{4.34}$$

Generating an Abridged Life Table: Estimating Values of $_nq_x$

More commonly demographers are interested not in collapsing complete life tables but in generating abridged life tables from scratch. As it was with the generation of single-year-of-age life tables, the key to this exercise is the calculation of probabilities of dying ($_nq_x$ values) from observed age-sex-specific death rates ($_nM_x$ values). By direct analogy with our earlier Eq. 4.1 we have:

$$_nM_x = ({}_nD_x/{}_nP_x)\,1{,}000 \tag{4.35}$$

Where $_nM_x$ denotes the death rate for males or females aged between exact ages x and x + n (i.e., aged x to x + n – 1 last birthday); $_nD_x$ = male or female deaths between exact ages x and x + n during the year for which the calculation is being performed (i.e., deaths at ages x to x + n – 1); $_nP_x$ = the mid-year male or female population aged between exact ages x and x + n (i.e., the mid-year population aged x to x + n – 1 last birthday).

We could also, if we wished, modify Eq. 4.35 in the same manner as we earlier modified Eq. 4.1 to obtain Eq. 4.2; i.e., we could use as our numerator the average annual number of deaths between exact ages x and x + n over the 3-year period straddling the reference year for which we require ${}_nM_x$ values. There is less reason to do this when constructing abridged life tables because random volatility in mortality data is reduced by the use of wider age intervals. That said, this ***is*** the approach taken in generating Table 4.5, simply because it was the approach taken in generating the equivalent Table 4.3. The indication that both life tables pertain to 2009–2011 says that the age-sex-specific death rates on which they are based relate average deaths over the 3-year period 2009–2011 to mid-year risk populations for the central year, 2010.

Equation 4.35 gives age-sex-specific death rates per 1,000 mid-year population at risk, and as was the case when using results from Eqs. 4.1 and 4.2 to obtain q_x values for the single-year-of-age life table, we ignore the multiplier of 1,000 in specifying ${}_nM_x$ values to be used in obtaining ${}_nq_x$ values for an abridged life table. Thus, ${}_nM_x = {}_nD_x / {}_nP_x$.

Recall Eqs. 4.7 and 4.8 earlier in the chapter, which gave values of q_x for the single-year-of-age life table from values of M_x. Analogous equations for an abridged life table are:

$$ {}_nq_x = (n.{}_nM_x) \,/\, \left(1 + {}^{n}\!/_{2}.{}_nM_x\right) \tag{4.36} $$

and:

$$ {}_nq_x = (2n.{}_nM_x) \,/\, (2 + n.{}_nM_x) \tag{4.37} $$

As with Eqs. 4.7 and 4.8, these are alternative forms of the same equation. They assume linear survivorship between exact ages x and x + n, and hence assume a separation factor of one-half. As previously noted, separation is an issue over all age intervals in an abridged life table. However, a frequently adopted approach (and the one taken in constructing Table 4.5) is to ***assume*** linear survivorship over 5-year age intervals between exact age 5 and the exact age marking the lower bound of the terminal open age interval (exact age 85 in Table 4.5). Error introduced by this approach is tolerable for most demographic purposes, and the appropriate version of Eq. 4.36 becomes:

$$ {}_5q_x = (5.{}_5M_x) \,/\, \left(1 + {}^{5}\!/_{2}.{}_5M_x\right) \tag{4.38} $$

Use of this equation to obtain ${}_nq_x$ values beyond exact age 5 (remember that $q_{85+} = 1.0$) still leaves us needing to find ${}_1q_0$ and ${}_4q_1$. The former can be calculated exactly as it was for the single-year-of-age life table (see the alternative Eqs. 4.9, 4.11 and 4.12 and associated discussion). In the case of ${}_4q_1$ an assumption of linear survivorship typically is unsound, and a separation factor of less than one-half is in order. The situation over this age interval is directly analogous to that during the first year of life. Just as infant deaths are concentrated at the younger end of the age interval bounded by exact ages 0 and 1, so, too, deaths between exact ages 1 and 5 tend to be concentrated at the younger end of that 4-year age interval.

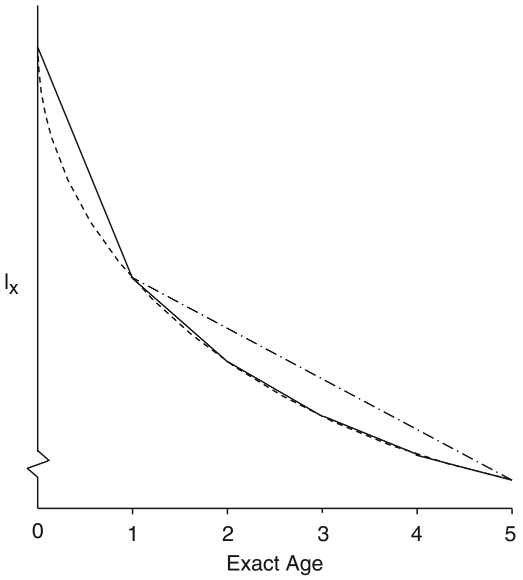


Fig. 4.5 Schematic illustration of how the survivorship function l_x approximates linearity over 1-year age intervals between exact ages 1 and 5, but does not do so over the combined 4-year interval or between exact ages 0 and 1

Figure 4.5 helps to explain the situation. It shows a schematic plot of the survivorship function l_x against age (the dashed curve), and straight line approximations of that function over successive one-year age intervals (the solid lines) and over the interval from exact age 1 to exact age 5 (the dot-dashed line). These straight lines represent linear survivorship over the age intervals to which they relate. The vertical axis is broken to signify that only the upper end of the full range of l_x values is portrayed. The point at which the survivorship curve meets the vertical axis corresponds with $l_x = l_0 = 100{,}000$.

You can see that between exact ages 0 and 1 the solid straight line is a poor approximation of the dashed curve. The survivorship function falls much more steeply early in this 1-year age interval than it does later in the interval. This is the situation invariably encountered with infant mortality, and the reason we introduced the concept of separation. An assumption of linear survivorship over this age interval amounts to an assumption that, on average, those who died in the interval did so at its mid-point or, to put it another way, survived for half a year. Clearly their average survival period was appreciably less than half a year, and we needed to adjust for this reality by using the separation factor f in calculating certain life table functions.

If we next focus on the age interval bounded by exact ages 1 and 5 we see that, when broken into 1-year segments, the series of four solid straight lines that traverse the interval in Fig. 4.5 collectively are quite a good approximation of the dashed survivorship curve. In other words, treating single years of age within the interval one by one (as the single-year-of-age life table does), an assumption of linear survivorship is not unreasonable. However, when we combine the four one-year age intervals into one four-year interval (as the abridged life table does) the straight line joining the points of intersection of the survivorship curve with exact ages 1 and 5 (the dot-dashed line) is again a poor approximation of the survivorship curve. We have the same sort of situation we have discussed at length with respect to infant mortality – deaths are concentrated at the younger end of the age interval, with the result that the survivorship function falls more steeply at that end of the interval. To assume linear survivorship through the interval (i.e., that those who die in the interval on average survive for half its width – 2 years) is not reasonable, and a separation factor with a value less than one-half is called for (because there are fewer deaths in the second half of the interval than in the first half).

It is helpful to pause at this juncture and re-conceptualize life table separation factors. When dealing with the separation factor used at age 0 we were able to conceive it as either (i) the proportion of deaths at age 0 in a birth cohort that occur in the year following the year of birth, or (ii) the proportion of deaths at age 0 in a calendar year that occur to children born the previous year. Because in an abridged life table we are no longer dealing with single-year age groups beyond exact age 1 an alternative conceptualization of the separation factor (equally applicable in a single-year age group context) is more readily understood. The separation factor may be thought of as ***the average proportion of the relevant age interval survived by persons who die in that age interval***. Chiang (1984: 141) refers to it as the ***'fraction of the last age interval of life'***.

Bearing this conceptualization in mind, the equation for ${}_4q_1$ may be specified as:

$$ {}_4q_1 = (4.{}_4M_1) / \left(1 + 4.\left(1 - f^*\right).{}_4M_1\right) \tag{4.39} $$

Where f* is a separation factor appropriate to mortality between exact ages 1 and 5.

This separation factor can be computed for any population for which a complete life table is available using a method outlined by Chiang (1984); the method makes use of values of q_x, p_x and separation factors for the four constituent single-year age groups that make up age group 1–4. The equation is:

$$ f^* = (q_1f_1 + p_1q_2\,(1{+}f_2) + p_1p_2q_3\,(2{+}f_3) + p_1p_2p_3q_4\,(3{+}f_4))\,/4\,(1 - p_1p_2p_3p_4) \tag{4.40} $$

Where q_x and p_x are standard single-year-of-age life table functions for $x = 1, 2, 3$ and 4; f_x values are separation factors for the same four single-year age groups. To date we have assumed them to all have a value of 0.50 (i.e., we have assumed linear survivorship across those age intervals), but Chiang (1984) suggests values of $f_1 = 0.43$, $f_2 = 0.45$, $f_3 = 0.47$ and $f_4 = 0.49$ are more accurate. These values reflect slight concentrations of mortality at the lower ends of the four single-year age groups.

Frequently, however, the desire is to use Eq. 4.39 in the construction of an abridged life table for a population for which no complete life table is available, so that Eq. 4.40 cannot be evaluated. In that circumstance it is common to use an estimate of f*. Fortunately values of f* generally lie in a fairly narrow range. As a guide to selecting an appropriate value Table 4.6 shows values of f* from

Table 4.6 Values of f* from Coale-Demeny and United Nations model life tables by sex and expectation of life at birth

	Expectation of life at birth							
Model	50	55	60	65	70	75	80	85
Male								
C-D East	0.33	0.33	0.32	0.34	0.36	0.38	0.38	0.38
C-D North	0.39	0.39	0.41	0.43	0.44	0.45	0.46	0.46
C-D South	0.31	0.31	0.33	0.35	0.36	0.38	0.39	0.40
C-D West	0.34	0.34	0.36	0.38	0.39	0.40	0.41	0.41
UN Chilean	0.34	0.34	0.34	0.36	0.38	0.39	0.40	0.41
UN Far Eastern	0.34	0.36	0.37	0.39	0.40	0.41	0.41	0.41
UN General	0.34	0.34	0.35	0.37	0.39	0.40	0.40	0.41
UN Latin American	0.34	0.34	0.35	0.37	0.38	0.39	0.40	0.41
UN South Asian	0.34	0.34	0.34	0.36	0.38	0.39	0.40	0.41
Female								
C-D East	0.33	0.33	0.31	0.32	0.33	0.34	0.35	0.35
C-D North	0.39	0.39	0.40	0.41	0.42	0.43	0.43	0.43
C-D South	0.31	0.31	0.33	0.34	0.35	0.35	0.36	0.37
C-D West	0.34	0.34	0.35	0.36	0.37	0.37	0.38	0.38
UN Chilean	0.34	0.34	0.34	0.35	0.36	0.36	0.37	0.38
UN Far Eastern	0.34	0.35	0.36	0.36	0.37	0.37	0.38	0.38
UN General	0.34	0.34	0.35	0.36	0.36	0.37	0.37	0.38
UN Latin American	0.34	0.34	0.35	0.35	0.36	0.37	0.37	0.38
UN South Asian	0.34	0.34	0.34	0.35	0.36	0.37	0.37	0.38

Coale-Demeny and United Nations model life tables by sex for expectations of life at birth ranging from 50 to 85 years. The Coale-Demeny model life tables (Coale and Demeny 1966; Coale et al. 1983) comprise an 'East' variant reflecting Eastern European populations, a 'North' variant based on Nordic populations, a 'South' variant based on Southern European populations and the 'West' variant previously noted to be based on populations from Western Europe, North America, Australasia, Japan, Taiwan and White South Africa. The United Nations model life tables for developing countries (United Nations 1982) offer 'Chilean', 'Far Eastern', 'Latin American' and 'South Asian' variants along with the 'General' variant that averages the other four. They are based on 36 pairs of male and female life tables for 10 Latin American, 11 Asian and one African population (Tunisia).

You can see from Table 4.6 (i) that f* values generally lie in the range 0.31 to 0.41 (higher for the Coale-Demeny North model, but this is an unusual model based on historical Scandinavian life tables), (ii) that they generally rise as expectation of life at birth increases, and (iii) that they tend to be higher for males, especially at higher life expectancies. A plausible estimate of f* for an abridged life table can be made having regard to the sex for which it is being constructed, an educated guess at the life expectancy at birth of the population in question, and an assessment of the most appropriate model to use. Thus, in constructing Table 4.5 above a value of

0.41 was used – it is a male life table for which e^o_0 has a value around 80, and the population is one for which the Coale-Demeny 'West' model is likely to be the most appropriate.

Note that f* is distinct from the separation factor f that we used in dealing with infant mortality. We cannot use the Keyfitz equation (Eq. 4.10) or the tables of separation factors presented as Tables 4.1 and 4.2 to obtain a separation factor for use in Eq. 4.39.

We can note at this point that in Eqs. 4.11, 4.39 and 4.38 we now have a family of equations for the various values of ${}_nq_x$ required to generate an abridged life table which satisfy the same general equation:

$$ {}_nq_x = (n.{}_nM_x) / (1 + n.(1 - F).{}_nM_x) \tag{4.41} $$

Where F is a separation factor relevant to mortality between exact ages x and x + n.

It remains the case, however, that problems may attach to the use of this equation to calculate ${}_1q_0$, and that equations requiring only registration data are preferable (see earlier discussion of Eq. 4.11). The correspondence between Eqs. 4.41 and 4.39 is clear, while that between Eqs. 4.41 and 4.38 is also clear when it is remembered that the latter assumes a separation factor of one-half.

You will sometimes (e.g., Preston et al. 2001) see Eq. 4.41 written as follows:

$$ {}_nq_x = (n.{}_nM_x) / (1 + (n - {}_na_x).{}_nM_x) \tag{4.42} $$

This introduces the function ${}_na_x$, which gives ***average person-years lived between exact ages x and x + n by persons dying in that age interval***. It is the product of the width of the age interval and the separation factor F relevant to mortality in that interval, or the interval width multiplied by the proportion of deaths in the interval that occur in the second half of the interval (i.e., ${}_na_x = n.F$, so that $F = {}_na_x/n$, which if substituted for F in Eq. 4.41 yields Eq. 4.42).

The foregoing discussion has provided a basic approach to estimating ${}_nq_x$ values for an abridged life table. Its chief drawback is assuming linear survivorship, and hence a separation factor of one-half, over 5-year age intervals beyond exact age 5. Other methods have been developed which take the issue of separation over these age intervals more seriously. Five will be discussed here – the Reed-Merrell method, the Greville method, the Fergany method, the Keyfitz-Frauenthal method and the method of reference to a standard life table.

The Reed-Merrell Approach to Estimating ${}_nq_x$

This method, once widely used but nowadays largely superseded, dates from an analysis of early U.S. life tables (Reed and Merrell 1939) that suggested there were equations that linked values of ${}_nq_x$ and ${}_nM_x$ across wide ranges of observations. While this 'traditional' method of life table construction (Preston et al. 2001: 45) has been criticized for its 'questionable generality' (Fergany 1971: 333) and being based on life tables whose widespread applicability there is 'no special reason to believe in' (Preston et al. 2001: 46), its equations are still sometimes used, especially to obtain values of ${}_5q_x$. It is thus desirable to be aware of them:

$$ {}_1q_0 = 1 - \exp\left[-{}_1M_0\left(0.9539 - 0.5509\,{}_1M_0\right)\right] \tag{4.43} $$

$$ {}_4q_1 = 1 - \exp\left[-4\,{}_4M_1\left(0.9806 - 2.079\,{}_4M_1\right)\right] \tag{4.44} $$

$$ {}_5q_x = 1 - \exp\left[-5\,{}_5M_x\left(1 + 0.2\,{}_5M_x\right)\right] \tag{4.45} $$

To the non-mathematical these equations will look horrendous. Equation 4.43 is almost never used and can be ignored, but the other two, especially Eq. 4.45, ***are*** encountered. Evaluating them manually should not prove difficult for competent users of a hand calculator offering the function 'e^x'. Simply evaluate the expression in square brackets following 'exp', hit 'e^x', and subtract the answer from 1. An alternative is to use Reed-Merrell conversion tables, which allow values of ${}_nq_x$ corresponding to values of ${}_nM_x$ to be read off. There are separate tables for ${}_1q_0$, ${}_4q_1$ and ${}_5q_x$, and also for ${}_1q_1$ and ${}_3q_2$ if an abridged table splits the interval between exact ages 1 and 5 into two – exact ages 1–2 and 2–5. There are separate equations analogous to Eqs. 4.43, 4.44 and 4.45 for ${}_1q_1$ and ${}_3q_2$ as well. However, the Reed-Merrell tables were prepared prior to the advent of electronic calculators and computers, and are nowadays superfluous. They can be found in Reed and Merrell (1939) or in appendices to Shryock, Siegel and Associates (1973) and Siegel and Swanson (2004). Note that Eqs. 4.43, 4.44 and 4.45 and the Reed-Merrell tables ***assume values of $_nM_x$ calculated to a base of 1.0***. Thus, if an age-specific death rate is specified as, say, 7.48 deaths ***per 1,000***, the equivalent ${}_nM_x$ is found by dividing by 1,000, or shifting the decimal point three places left (i.e., ${}_nM_x = 0.00748$).

The Greville Approach to Estimating $_nq_x$

This approach to dealing with non-linear survivorship between exact ages specified in an abridged life table (Greville 1943) focuses on the general equation for ${}_nq_x$ presented above as Eq. 4.41, and provides a further equation for obtaining the appropriate separation factor F for substitution in this equation. The Greville equation for the separation factor is:

$$ F = {}^1\!/_2 - \left[n\left({}_nM_x - 0.095\right)\right]/12 \tag{4.46} $$

To obtain a value of ${}_nq_x$ corresponding to an observed value of ${}_nM_x$ using Greville's approach, evaluate Eq. 4.46 for the observed ${}_nM_x$, then substitute back in Eq. 4.41 for both F and ${}_nM_x$ to obtain ${}_nq_x$. Although this approach is sometimes claimed to

Table 4.7 Abridged life table for Australian males, 2009–2011 constructed using Greville separation factors beyond exact age 5

Age (x)	n	$_nM_x$	$_nF_x$	$_nq_x$	l_x	$_nd_x$	$_nL_x$	T_x	e^o_x
0	1	0.00498	0.11500	0.00496	100000	496	99561	8019757	80.20
1	4	0.00021	0.41000	0.00082	99504	82	397722	7920196	79.60
5	5	0.00012	0.53953	0.00058	99422	58	497348	7522474	75.66
10	5	0.00011	0.53953	0.00056	99364	56	496688	7025126	70.70
15	5	0.00046	0.53939	0.00230	99308	229	496013	6528438	65.74
20	5	0.00062	0.53933	0.00309	99079	306	494690	6032425	60.89
25	5	0.00074	0.53928	0.00367	98773	362	493030	5537735	56.07
30	5	0.00094	0.53919	0.00469	98411	462	490991	5044705	51.26
35	5	0.00118	0.53909	0.00587	97949	575	488420	4553714	46.49
40	5	0.00158	0.53892	0.00787	97374	767	485103	4065294	41.75
45	5	0.00229	0.53863	0.01140	96607	1102	480494	3580191	37.06
50	5	0.00344	0.53815	0.01710	95505	1633	473754	3099697	32.46
55	5	0.00520	0.53742	0.02570	93872	2413	463779	2625943	27.97
60	5	0.00800	0.53625	0.03926	91459	3591	448968	2162164	23.64
65	5	0.01303	0.53415	0.06327	87868	5559	426392	1713196	19.50
70	5	0.02179	0.53050	0.10364	82309	8531	391520	1286804	15.63
75	5	0.03776	0.52385	0.17322	73778	12780	338465	895284	12.13
80	5	0.06758	0.51142	0.29003	60998	17691	261772	556819	9.13
85	ω	0.14678		1.00000	43307	43307	295047	295047	6.81

be appropriate for all age intervals in an abridged life table ***it is not recommended for the calculation of*** $_1q_0$ ***or*** $_4q_1$, because it is based on a model of mortality that is unsound at those ages. It is best to use standard procedures at these younger ages, and to treat Greville's equation as a useful method of allowing for non-linear survivorship in generating $_nq_x$ values beyond exact age 5.

Table 4.7 shows the abridged life table presented above as Table 4.5 reconstructed using Greville separation factors beyond exact age 5. In other words Eqs. 4.46 and 4.41 have been used to calculate $_5q_x$ values, where in Table 4.5 Eq. 4.38 was used. Separation factors have been indicated in a new column ($_nF_x$). The first separation factor ($_1F_0$) is the one calculated in Table 4.4, while the second ($_4F_1$) is the separation factor for that age interval from the Coale-Demeny male 'West' model life table for $e^o_0 = 80$ extracted from Table 4.6. There are differences between the final columns of Tables 4.5 and 4.7, but they are minor. Allowing for non-linear survivorship at the oldest ages (the reality that those dying in these age groups on average survive for a bit longer than 2.5 years) adds marginally to life expectancies at those older ages, and these increments feed through to slightly higher life expectancies at younger ages and at birth in Table 4.7 as well.

The Fergany Method of Estimating ${}_nq_x$

A deceptively simple approach to converting ${}_nM_x$ values to life table ${}_nq_x$ values was suggested by Fergany (1971). He was reacting to perceived deficiencies in the Reed-Merrell and Greville approaches, both of which he claimed (1971: 333) ‘involve assumptions that are either unjustifiable or of questionable generality and rough approximations’. He was also unimpressed with a method suggested by Chiang (1968) because of its ‘exacting data requirements’. The Fergany conversion equation is:

$$ {}_nq_x = 1 - \exp\left(-n.{}_nM_x\right) \tag{4.47} $$

Reed and Merrell had actually derived this equation and discarded it as unsatisfactory for transforming age-specific death rates to probabilities of death, Fergany (1971: 333) describing it as ‘very strange indeed that their justification for this went unquestioned by demographers for more than thirty years.’ Essentially that justification was a perceived poor fit to 1910 US life tables, but these had been derived using ‘a host of … intricate actuarial techniques’ that made them dubious models against which to assess the merit of a ‘universal’ technique for transforming age-specific death rates into life table probabilities of death. Fergany saw the simplicity of his approach as a major attribute – it required no special conversion tables, did not involve a complicated formula, and had no parameters that needed to be estimated empirically or borrowed from a population other than the one the life table was being constructed for. All it needed was age-specific death rates and, at the time, widely available exponential function tables. The latter have since been superseded by electronic calculators and computers.

The Keyfitz-Frauenthal Method of Estimating ${}_nq_x$

This approach to converting ${}_nM_x$ values into life table ${}_nq_x$ values has been shown by Keyfitz and Frauenthal (1975) to embrace the Reed-Merrell and Greville methods as special cases. It provides the following equation for converting age-specific death rates to ${}_nq_x$ values for an abridged life table:

$$ {}_nq_x = 1 - \exp\left[-n.\left\{{}_nM_x + \left({}_nP_{x-n} - {}_nP_{x+n}\right)\ \left({}_nM_{x+n} - {}_nM_{x-n}\right)/48{}_nP_x\right\}\right] \tag{4.48} $$

Where ${}_nP_x$ is the observed population aged between exact ages x and x + n (i.e., it is the denominator of ${}_nM_x$ – see Eq. 4.35).

You will note that Eq. 4.48 makes use of values of P and M for n-year age intervals immediately below and above the n-year interval for which a value of ${}_nq_x$ is being calculated. This implies that it cannot be used to obtain ${}_1q_0$ and ${}_4q_1$, as there are no 1-year and 4-year age intervals, respectively, below the age intervals to which these probabilities of dying apply. Questions also arise with respect to the equation's application to obtain ${}_5q_5$ and ${}_5q_{80}$. In the former instance we need to combine the 1-year and 4-year intervals that typically begin an abridged life table to yield a 5-year interval below the one of interest. In the latter we would need data for the interval between exact ages 85 and 90, which is part of the terminal ***open*** interval in the standard abridged life table format, in order to have a 5-year interval above the one of interest.

Using Reference to a Standard Life Table to Estimate ${}_nq_x$

A further approach to the conversion of observed ${}_nM_x$ values into life table ${}_nq_x$ values bases it on the relationship that exists between these values in an existing, ***standard*** life table. The standard life table may be either a single-year-of-age life table or an abridged life table, with the proviso that it cannot be ***more*** abridged than the life table one wishes to construct. It should be selected so that the levels of mortality by age (the patterns of mortality) in the standard population and the population of interest (as measured by observed age-specific death rates ${}_nM_x$) are roughly comparable. It would not, for example, be appropriate to use a life table for Australian males in 2009–2011 as the standard to construct a life table for males in a developing African country around the same date; the patterns of mortality would be very different. It is also usual to employ a standard life table for the same sex as the population for which a life table is being constructed, since mortality patterns often differ significantly by sex.

Frequently, judicious use is made of model life tables (United Nations 1955; Coale and Demeny 1966; Brass 1971; United Nations 1982; Coale et al. 1983; Coale and Guo 1989; Murray et al. 2003; Zhao 2007) to provide appropriate standards. As previously explained, model life tables generalize mortality patterns found in ***families*** of actual life tables, the skill in using them lying in selecting the 'family' to which mortality conditions in a population of interest are most likely to belong. A major initial reason for developing model life tables was to assist in analysing the mortality of developing populations, for which reliable mortality data were often lacking. In particular they were often used to gain insight into likely adult mortality conditions in such populations, on the basis of more reliable empirical evidence about infant and childhood mortality. More recent developments have sought to address such things as the very low levels of mortality now prevalent in many developed countries, which did not exist when the United Nations Population Division, Coale and Demeny, and Brass were pioneering the field, and the huge impact of HIV/AIDS on mortality patterns in African and some Asian nations in particular.

To use the method of reference to a standard life table we compute a quantity ${}_ng_x$ using the equation:

$${}_ng_x = \left(n/{}_nq_x^s\right) - \left(1/{}_nM_x^s\right) \tag{4.49}$$

Where ${}_n g_x$ actually represents the average number of person-years of life lost between exact ages x and x + n by persons dying in that age interval in the standard population; ${}_n q_x^s$ and ${}_n M_x^s$ are the ${}_n q_x$ and ${}_n M_x$ values for the age interval bounded by exact ages x and x + n for the ***standard*** population (assuming that ${}_n M_x^s = {}_n m_x^s$).

Having evaluated Eq. 4.49, ${}_n g_x$ is substituted back in the following equation for ${}_n q_x$:

$$ {}_n q_x = n.{}_n M_x / \left(1 + {}_n g_x . {}_n M_x\right) \tag{4.50}$$

As an aside, comparison of Eq. 4.50 with the generalized equation for ${}_n q_x$, Eq. 4.41, indicates that ${}_n g_x = n\,(1 - F)$ (where F is the separation factor appropriate to the age interval in question). It is from this relationship that the interpretation of ${}_n g_x$ given above derives. Where deaths are concentrated at the lower end of an age interval, those dying in the interval obviously survive, on average, for less than half the interval; or putting it another way, they ***fail*** to survive for ***more*** than half the interval. Since we noted earlier that when deaths are concentrated at the lower end of an age interval $F < {}^1/_2$, it follows that ${}_n g_x = n\,(1 - F) > {}^n/_2$, which is consistent with ${}_n g_x$ measuring average ***failure*** to survive, or ***life lost***. Similarly, if deaths are concentrated at the upper end of an age interval those who die in the interval on average survive for more than half the interval and ***fail*** to survive for ***less*** than half of it. Since in this circumstance $F > {}^1/_2$, ${}_n g_x = n\,(1 - F) < {}^n/_2$, which again is consistent with ${}_n g_x$ measuring average ***failure*** to survive, or ***life lost***.

Equation 4.50 can also be compared with Eq. 4.42 to reveal that ${}_n g_x = n - {}_n a_x$. So the average number of person-years of life ***lost*** between exact ages x and x + n by persons dying in that age interval (${}_n g_x$) is equal to the width of the interval (n) less the average number of person-years of life ***lived*** between exact ages x and x + n by persons dying in that age interval (${}_n a_x$). We can use this relationship to develop an alternative approach to obtaining ${}_n q_x$ values by reference to a standard life table. Substituting ($n - {}_n a_x$) for ${}_n g_x$ in Eq. 4.49 yields:

$$ {}_n a_x = n - \left(n/{}_n q_x^s\right) + \left(1/{}_n M_x^s\right) \tag{4.51}$$

After evaluating this equation we can substitute for ${}_n a_x$ in Eq. 4.42 to obtain ${}_n q_x$.

Other schemes for converting observed ${}_n M_x$ values into life table ${}_n q_x$ values also exist in the literature, and there have in addition been techniques developed for ***graduating***, or ***smoothing***, series of ${}_n M_x$ values prior to translation to ${}_n q_x$ values. This sort of technical detail, however, need not concern us in this introductory text, the aim of which is to provide a basic coverage of life table construction. All manner of methodological refinements are discussed in the literature, but many are of more concern to actuaries than to demographers.

You might, however, reasonably ask when, assuming that equations which assume linear survivorship have been ruled out, each of the approaches to obtaining

${}_nq_x$ values just outlined should be used. There are no hard and fast rules. Differences between results the methods produce ordinarily are not large (e.g., both Fergany (1971) and Keyfitz and Frauenthal (1975) show their methods to yield similar results to the methods of Reed and Merrell, and Greville), and when dealing with populations of developing countries, data deficiencies leading to unreliable ${}_nM_x$ values will probably have far greater bearing on the accuracy of a life table than will the choice of a method of life table construction. The method of reference to a standard life table can be especially convenient if, for example, there is a desire to construct a life table at some other date for a population for which a life table already exists to act as a standard.

Generating an Abridged Life Table: Estimating Other Life Table Functions

Once a set of ${}_nq_x$ values have been calculated for an abridged life table, the remainder of the table can be constructed using the following relationships and the fact that $l_0 = 100{,}000$.

$$l_{x+n} = l_x - l_x.{}_nq_x = l_x\,(1 - {}_nq_x) \tag{4.52}$$

$${}_nd_x = l_x.{}_nq_x = l_x - l_{x+n} \tag{4.53}$$

$${}_np_x = 1 - {}_nq_x \tag{4.54}$$

and, after calculating ${}_1L_0$ using Eq. 4.23:

$${}_4L_1 = f^*.l_1 + \left(4 - f^*\right).l_5 \tag{4.55}$$

and, for $x \geq 5$, assuming linear survivorship:

$${}_nL_x = {}^n\!/_2\,(l_x + l_{x+n}) \tag{4.56}$$

or, for $x \geq 5$, assuming ${}_nm_x = {}_nM_x$:

$${}_nL_x = {}_nd_x / {}_nM_x \tag{4.57}$$

and, for the last row of the life table:

$$L_{x+} = T_x = d_{x+}/M_{x+} = l_x/M_{x+} \tag{4.58}$$

then, for each row before the last, working up the life table:

$$T_x = T_{x+n} + {}_nL_x \tag{4.59}$$

and finally:

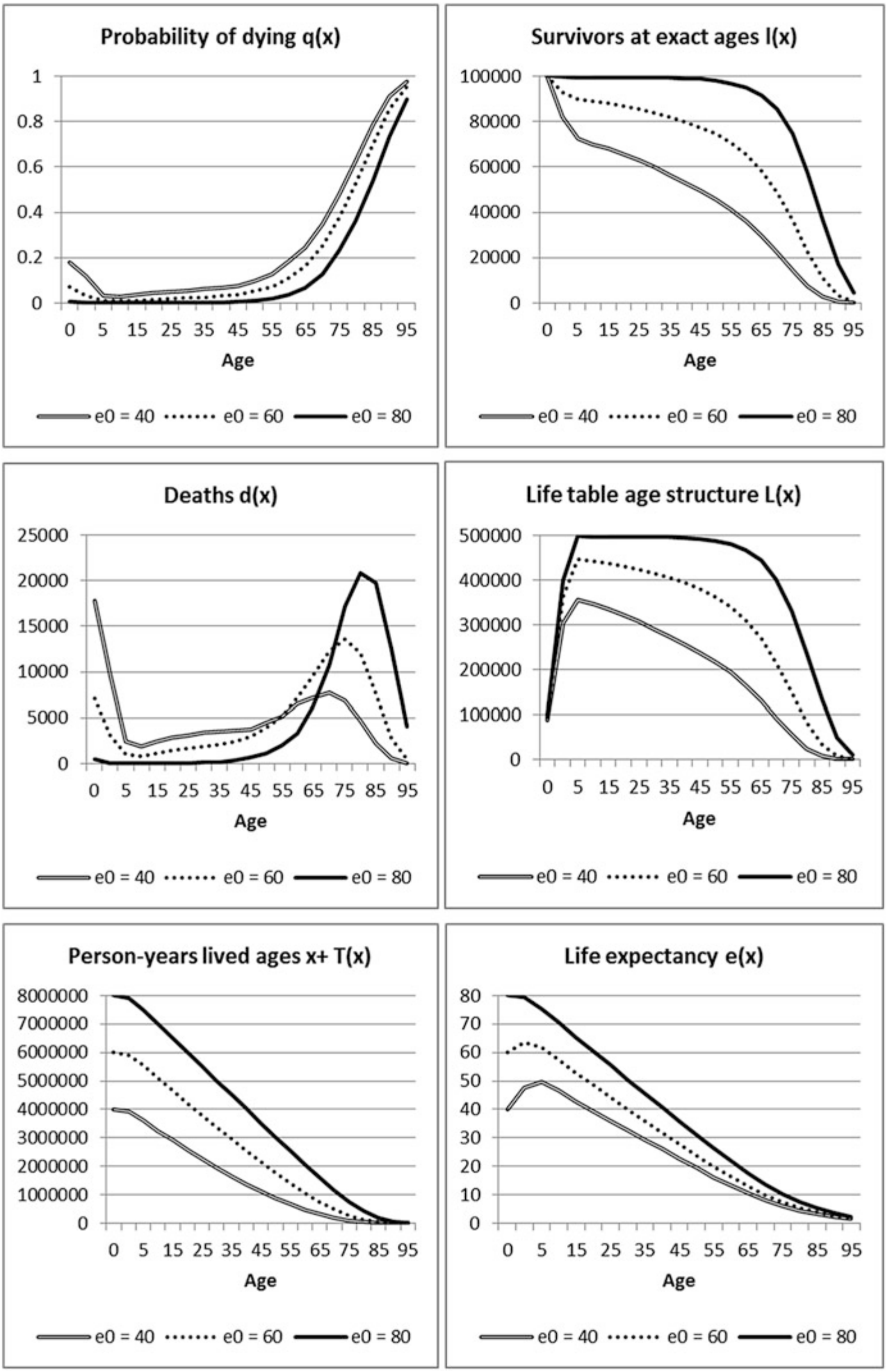
$$e^o{}_x = T_x/l_x \tag{4.60}$$

The Reed-Merrell and Keyfitz-Frauenthal approaches to the construction of abridged life tables and the method of reference to a standard life table do, in addition to providing procedures for obtaining values of ${}_nq_x$, provide special procedures for calculating the ${}_nL_x$ and T_x columns. However, they are not widely used and will not be elaborated on here. The interested reader is referred to Siegel and Swanson (2004: 312–315). The equations above are satisfactory for present purposes, but note that Eq. 4.56 should be used in conjunction with Eq. 4.38, while Eq. 4.57 is appropriate when values of ${}_nq_x$ have been obtained by a method which does not assume linear survivorship between exact ages x and x + n.

The Shapes of Life Table Functions

Having now completed discussion of the mechanics of constructing single-year-of-age and abridged life tables it is appropriate to pause and examine the shapes of life table functions and how these vary between high mortality (low life expectancy) and low mortality (high life expectancy) populations. Figure 4.6 shows plots of several life table functions from Coale and Demeny 'West' model life tables for female populations with life expectancies at birth of 40, 60 and 80 years. With respect to the mortality transition (a component of the familiar 'demographic' transition) from high to low mortality these life tables can be regarded as respectively pre-transitional, mid-transitional and post-transitional, or tables pertaining to high, intermediate and low mortality populations. Function plots derive from abridged life tables which recognise age groups 0, 1–4, then 5-year age groups 5–9 to 95–99. The open interval 100+ is excluded.

Life table probabilities of dying (${}_nq_x$ values) derived from age-specific death rates (${}_nM_x$ values) are much higher at ages 0 and 1–4 for the high mortality population than for the low mortality population, with the mid-transitional population falling in between. It is typical of higher mortality populations that they have much higher rates of infant and early childhood mortality (i.e., under-5 mortality) than do low mortality populations, and this reality is a major factor in the differential life expectancies at birth they exhibit, because in higher mortality populations so many members forfeit (through premature death) nearly all of their ***potential*** contribution to the T_x column of the life table. At older ages ${}_nq_x$ values rise earlier and to ultimately higher levels for the low and intermediate life expectancies than for life expectancy at birth = 80 years, with values for the latter, post-transitional, population only beginning to move noticeably above zero in middle age (late 40s and 50s).



Probability of dying q(x)

Survivors at exact ages l(x)

Deaths d(x)

Life table age structure L(x)

Person-years lived ages x+ T(x)

Life expectancy e(x)

Fig. 4.6 Graphs of life table functions for west model abridged life tables: Female life expectancies at birth of 40, 60 and 80 years

Survivorship (l_x) functions for the three populations are very different. Under conditions of high mortality there is an initial steep fall over the infant and early childhood ages (a loss of almost 18 and almost 28 % of the radix population by exact ages 1 and 5 respectively), a noticeable though less steep decline through the late childhood and early and middle adult ages, then a steepening again beyond age group 50–54 until sheer lack of survivors intervenes at the very oldest ages. By contrast under conditions of low mortality the survivorship curve is flat to about ages 50–54, and only then begins to fall increasingly rapidly. It is sobering in comparing these l_x plots to realize that the declines in l_x over the first year of life for $e^o_0 = 40$ in Fig. 4.6 are not equalled for $e^o_0 = 80$ until beyond exact age 70. This should strongly reinforce in your mind the huge potential that initiatives to reduce infant mortality have for improving longevity in high mortality populations, and indeed their key importance historically in bringing about mortality transitions across the globe. Once again the mid-transitional population follows a path intermediate between those of the high and low mortality populations.

The importance of infant and early childhood mortality in high mortality populations is again clear in the comparative plots of ${}_nd_x$ in Fig. 4.6. Very large numbers of deaths occur at the youngest ages in a high mortality population. By contrast very few occur before about age 40 in low mortality populations, where deaths peak at ages in the late 70s and 80s. For the high mortality population there is a secondary peak at ages 70–74, but it is nowhere near as pronounced as the infant mortality peak, and we are comparing here a 5-year age group with a 1-year age group. The ${}_nd_x$ plot for the mid-transition population is once again intermediate. There are still appreciable numbers of deaths in infancy and early childhood for this population, but the peak at older ages is more pronounced and located at ages 75–79, between the peaks for the high and low mortality populations.

Age structures for the three life table stationary populations are depicted by the ${}_nL_x$ plots. The comparatively larger size of the life table stationary population as life expectancy at birth increases is clear, and while the high mortality population loses a lot of members very young its stationary population also has a comparatively young age structure due to the cumulative effect with increasing age of having higher probabilities of dying at every age. Its value of ${}_5L_{80}$, for example, the number aged 80–84 in the life table stationary population, is barely one-tenth of that for the low mortality post-transitional population.

Unsurprisingly the graphs of person-years lived by the life table stationary population beyond exact age x in Fig. 4.6, represented by the T_x columns of the respective life tables, show much larger numbers at each age in the low mortality population than in the high mortality population, although at the very oldest ages numbers in all three populations naturally converge toward zero. The nature of the T_x function of course dictates that all trend lines fall progressively from the youngest to the oldest ages.

Aside from life expectancies being higher, age for age, in the low mortality population than in the high mortality population, with those for the mid-transitional population in between, the most interesting difference between graphs of e^o_x in Fig. 4.6 is that for the high mortality and mid-transitional populations life expectancy ***increases*** from its level at birth for survivors at exact age 1 and (for the high mortality population) exact age 5. This is a common feature of mortality conditions in higher mortality populations. Those who manage to survive through the high mortality infant and early childhood age groups actually ***improve*** their subsequent survival chances by doing so.

Using Life Tables: Survival

How do demographers use life tables? In a variety of ways, but a number of very important applications hinge on measures known as ***life table survival ratios***. They are used in such exercises as making population projections, estimating net intercensal migration, calculating the net reproduction rate (NRR), checking the accuracy of age statement and completeness of enumeration at a census, checking the completeness of birth registration, and estimating fertility from a population's age distribution. ***A survival ratio expresses the proportion of persons who reach a given age alive who survive to attain a nominated subsequent age***. These ages may be exact ages or ages in completed years (i.e., age groups).

Population Survival Ratios

Survival ratios need not necessarily derive from life tables. Indeed many useful applications of life tables entail equating a life table survival ratio with an equivalent survival ratio derived from census and/or vital registration data in which one element (numerator or denominator) is unknown, then solving for that unknown. ***Population survival ratios, usually based on census data, give the ratio of survivors in an age group at some later time to the population in the equivalent younger age group at some earlier time***. Thus they measure the proportion surviving, or the probability of survival, from the younger to the older age group.

For a single-year age group x, the population survival ratio over an n-year period from age x to age x + n is given by:

$$\mathrm{PSR}_{\mathrm{x,x+n}} = \mathrm{P(y+n)_{x+n}/P(y)_x} \tag{4.61}$$

Where $\mathrm{P(y)_x}$ is the population aged x in year y; n is the interval in years over which the population is being survived.

For an age group x to x + t − 1 (where t is the width of the age group in years) the population survival ratio over an n-year period is given by:

$$\mathrm{PSR}_{\mathrm{x\ to\ x+t-1,x+n\ to\ x+t-1+n}} = \mathrm{P(y+n)_{x+n\ to\ x+t-1+n}/P(y)_{x\ to\ x+t-1}} \tag{4.62}$$

Where $P(y)_{x\ \text{to}\ x+t-1}$ is the population aged x to $x + t - 1$ in year y; n is once again the interval in years over which the population is being survived.

In case specifying an age group as 'x to $x + t - 1$' is a source of confusion, think of the manner in which we conventionally specify 5-year age groups – 0–4 years, 5–9 years, 10–14 years, etc. Although these are ***5***-year age groups, because we specify the upper and lower limits in terms of ages ***in completed years*** rather than in terms of ***exact*** ages, the difference between the upper and lower limits is four, or ***one less*** than the true age group width. The '-1' in '$x + t - 1$' simply takes account of this in, following convention, specifying the upper limit of an age group as an age in completed years rather than an exact age.

Note that in Eqs. 4.61 and 4.62 the numerators refer to ***survivors*** at the later point in time (year $y + n$) from the populations forming the denominators. This implies that ideally one should make allowance for ***migration*** over the n-year survival period; numerators should be adjusted upward or downward from straight population figures to take account of net out-migration or net in-migration respectively. Having made this point we will proceed for the moment as if dealing with a closed population. The migration factor is, however, important, to the point where a major application of population survival ratios is the ***estimation*** of net (generally intercensal) migration.

Note also that when we refer in Eqs. 4.61 and 4.62 to a population in year y and an equivalent population in year $y + n$ ***it is assumed that those populations refer to the same date in each year (or very close to it)***. If there is some significant departure from this assumption (as, for example, when we find that successive censuses were held on different dates during the year) the value of n, the survival interval, becomes something other than a whole number of years and calculations are not quite so straightforward. One approach to this problem is to adjust one of the sets of census data to the same date during the year as applies to the other set, ***then*** calculate survival ratios.

Note further that population survival ratios are ***multiplicable***. The survival ratio over any age range is equal to ***the product of survival ratios over any number of discrete component age ranges*** (i.e., age ranges which do not overlap, but which together exactly make up the larger age range). For example:

$$PSR_{0,5} = PSR_{0,1}\ .\ PSR_{1,5}$$

since:

$$P(y+5)_5/P(y)_0 = (P(y+1)_1/P(y)_0)\,.\,(P(y+5)_5/P(y+1)_1)$$

Similarly:

$$PSR_{0-4,10-14} = PSR_{0-4,5-9}\ .\ PSR_{5-9,10-14}$$

since:

$$P(y + 10)_{10-14}/P(y)_{0-4} = (P(y + 5)_{5-9}/P(y)_{0-4}) \cdot (P(y + 10)_{10-14}/P(y + 5)_{5-9})$$

In each case elements on the righthand sides of these expressions cancel. These are straightforward examples, with the survival age ranges split into just two components, but the multiplicative principle extends to splits into larger numbers of components.

Survival Ratios from Birth

Survival ratios from birth are a variation on population survival ratios. They measure ***the ratio of survivors in an age group at some later time to the size of the cohort forming that age group at birth***. In other words they measure the proportion of the birth cohort who survive to form the age group, or the probability of survival from birth to the age group.

The distinction between survival ratios from birth and population survival ratios can best be understood using the Lexis diagram. Population survival ratios pertain to situations of the type illustrated by Fig. 4.7a (note that both Lexis diagrams in Fig. 4.7 feature grid squares with dimensions 5 years of age by 5 years of calendar time, rather than the more familiar one year by one year squares). This diagram depicts survival from age group 0–4 years at a census held on 30:6:96 to age group 10–14 at a census held 10 years later on 30:6:06. This is an example of survival between ***age groups*** (populations we represent on a Lexis diagram by vertical lines), and that is what population survival ratios measure.

Figure 4.7b, on the other hand, illustrates survival from birth to the age group 0–4 years at a census held on 30:6:06. Age group 0–4 years is a 5-year age group, and in a closed population clearly comprises survivors from births over the 5-year period immediately preceding the census – that is, births between 1:7:01 and 30:6:06. These births constitute a population attaining exact age 0, which we represent on a Lexis diagram by a horizontal line. The population (age group) to which this cohort is being survived, however, is one we represent on a Lexis diagram by a vertical line. Thus in Lexis diagram terms survival from birth is concerned with survival from a population we represent by a horizontal line at exact age 0 to one we represent by a vertical line at some later point up the cohort diagonal band which the extremities (the end points) of the former line define.

The survival ratio ***from birth to a single-year age group x*** is given by:

$$SR_{B,x} = P(y)_x/B\,(y - x - 1) \tag{4.63}$$

Where $P(y)_x$ is the population aged x at a specified date in year y; $B(y - x - 1)$ is births during the year commenced the date of the year y population count in year $y - x - 1$.

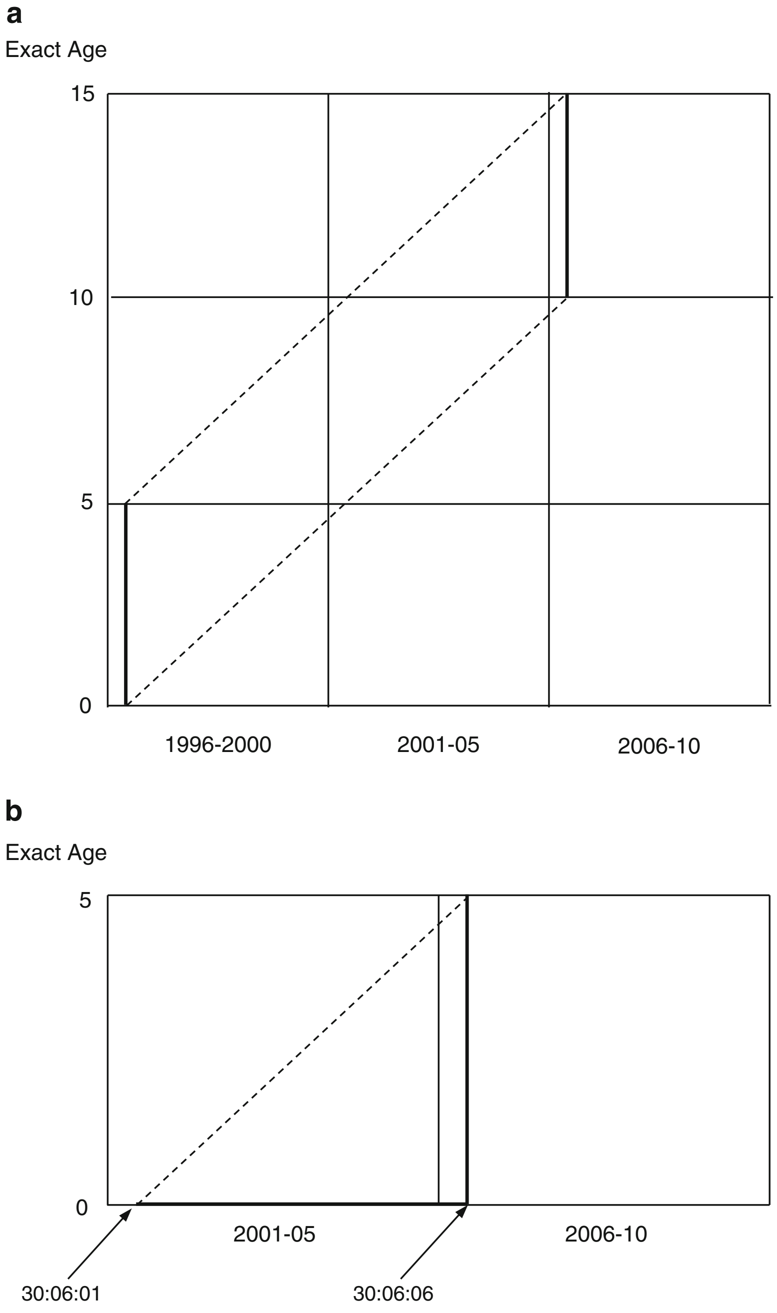


Fig. 4.7 Lexis diagram illustrations of survival between age groups and survival from birth (**a**) Survival from age group 0–4 in mid-1996 to age group 10–14 in mid-2006 (**b**) Survival from birth to age group 0–4 in mid-2006

The survival ratio ***from birth to an age group x to x + t − 1*** (where t is the width of the age group in years) is given by:

$$SR_{B,x\ \mathrm{to}\ x+t-1} = P(y)_{x\ \mathrm{to}\ x+t-1}/B^{t,y-x-t} \tag{4.64}$$

Where $y =$ year of the population count which includes survivors aged x to $x + t - 1$; $x =$ lower limit of age group survived to; $P(y)_{x\ \mathrm{to}\ x+t-1}$ is the population aged x to $x + t - 1$ in year y; $B^{t,y-x-t}$ is ***births during a t-year period commenced the date of the year y population count in year y − x − t***.

Taking the example illustrated in Fig. 4.7b, $y = 2006$, $x = 0$ and $t = 5$. Therefore:

$$SR_{B,0-4} = P(2006)_{0-4}/B^{5,2001}$$

In other words the survival ratio from birth to age group 0–4 is the ratio of the population aged 0–4 at the date of the 2006 census to the births that occurred over a 5-year period commenced midnight on 30th June (the date of the 2006 census) in 2001.

Life Table Survival Ratios

The true value of population survival ratios and survival ratios from birth becomes apparent when they are linked with ***life table*** survival ratios. There are several important applications of this sort of strategy, which basically involve:

1. Equating a population survival ratio or a survival ratio from birth with an equivalent life table survival ratio (derived from a life table that summarizes the mortality experience of the population in question during the period over which it is being studied).
2. Using this equality to obtain one of the quantities (numerator or denominator) from the population survival ratio or the survival ratio from birth ***which is unknown***.

In other words we don't actually calculate the population survival ratio or the survival ratio from birth, because ***we can't*** – we don't know one of the quantities needed. Instead we obtain an ***expected*** value of the survival ratio from a life table and use this to estimate the unknown quantity (either a population in some age group or a number of births over a defined period).

Recall that the life table population is a ***stationary population***, the age structure of which is given by the L_x column. Use of this column enables us to calculate life table survival ratios which express the probability of surviving from any age to any later age under the mortality conditions summarized by the life table. In general, for a single-year-of-age life table:

$$LTSR_{x,x+n} = L_{x+n}/L_x \tag{4.65}$$

Where $LTSR_{x,x+n}$ means the life table survival ratio from age x to age x + n.

Like population survival ratios, life table survival ratios are ***multiplicable***; the life table survival ratio over any age range is equal to the product of life table survival ratios over any number of discrete component age ranges. Thus, for example:

$$LTSR_{0,5} = LTSR_{0,1} \; . \; LTSR_{1,5}$$

The life table survival ratio from age 0 to age 5 is the product of the life table survival ratios from age 0 to age 1 and from age 1 to age 5.

Life table survival ratios may also be computed from abridged life tables. In general:

$$LTSR_{x \text{ to } x+t-1, x+n \text{ to } x+t-1+n} = {}_tL_{x+n}/{}_tL_x \tag{4.66}$$

Where t = width of age group for which the life table survival ratio is being calculated; n = number of years over which the life table population aged x to x + t − 1 is survived.

So, for example:

$$LTSR_{0-4,10-14} = {}_5L_{10}/{}_5L_0 = {}_5L_{10}/\left({}_1L_0 + {}_4L_1\right)$$

Since ${}_5L_0 = {}_1L_0 + {}_4L_1$, because an abridged life table begins with a 1-year, then a 4-year age interval.

The life table survival ratios discussed to this point have been equivalents of ***population*** survival ratios; they have measured survival from one age ***group*** in the life table stationary population to another. We can also calculate life table survival ratios which measure survival from a given ***exact*** age to an older age ***group***. In particular, we can calculate the life table equivalent of a survival ratio ***from birth***; a ratio measuring the probability of surviving from exact age 0 to some subsequent age group.

For survival from birth to a single-year age group x we have:

$$LTSR_{B,x} = L_x/l_0 \tag{4.67}$$

For survival from birth to an age group of width t years we have:

$$LTSR_{B,x \text{ to } x+t-1} = {}_tL_x/t.l_0 \tag{4.68}$$

Where t = width in years of the age group survived to; x = lower limit of age group survived to.

So, for example:

$$\mathrm{LTSR}_{\mathrm{B},15-19} = {}_5\mathrm{L}_{15}/5\ .\ \mathrm{l}_0$$

Why do we need to multiply l_0 in the denominator by t? Because the population in any t-year age group are survivors from births over a t-year period. And, of course, the life table stationary population experiences l_0 births ***each year***. The population aged 0–4, for example, comprises survivors from ***five*** single-year (though not necessarily ***calendar*** year) birth cohorts.

Note that ***when calculating a life table survival ratio a numerator or denominator is obtained from the $_n L_x$ column of the life table if it refers to an age group, and from the l_x column if it refers to an exact age***. If confused, try to think of your problem in Lexis diagram terms; if your survival ratio numerator or denominator is a population you would represent on a Lexis diagram by a vertical line, use the L_x column; if it is a population you would represent by a horizontal line, use the l_x column. ***There is one modification to this principle*** which we make for practical reasons. If an age group of interest is of the type 'ages x ***and older***', while we ***could*** use the $_\mathrm{n}\mathrm{L}_\mathrm{x}$ column in accordance with our principle (summing all values in the column from exact age x to the end of the life table), ***it is simpler to consult the T_x column***, which gives precisely this summation directly. Thus the life table survival ratio from ages x ***and older*** to ages x + n ***and older*** is simply $\mathrm{T}_{\mathrm{x+n}}/\mathrm{T}_\mathrm{x}$.

Relationship Between Population and Life Table Survival Ratios

If migration is zero during a period, then a population survival ratio equals the equivalent life table survival ratio; that is, the equivalent survival ratio derived from a life table representative of the mortality of the population in question over the period in question. Thus, from our earlier equations for a PSR and a LTSR from age x to age x + n (Eqs. 4.61 and 4.65) we have:

$$\mathrm{P(y+n)_{x+n}/P(y)_x = L_{x+n}/L_x}$$

It follows that:

$$\mathrm{P(y+n)_{x+n} = P(y)_x.\,(L_{x+n}/L_x)} \qquad (4.69)$$

Here we are obtaining the population in a given single-year age group at some ***later*** point in time as the product of the population in an equivalent ***younger*** age group at an ***earlier*** point in time and an appropriate life table survival ratio. This procedure is known as ***FORWARD SURVIVAL***. Equation 4.69 is one for use when forward surviving single-year age groups.

Similarly, from our earlier equations for a PSR and a LTSR from age group x to $x+t-1$ to age group $x+n$ to $x+t-1+n$ (i.e., for an age group of width t survived over a period of n years – see Eqs. 4.62 and 4.66) we have:

$$P(y+n)_{x+n \text{ to } x+t-1+n}/P(y)_{x \text{ to } x+t-1} = {}_tL_{x+n}/{}_tL_x$$

It follows that:

$$P(y+n)_{x+n \text{ to } x+t-1+n} = P(y)_{x \text{ to } x+t-1} \cdot ({}_tL_{x+n}/{}_tL_x) \tag{4.70}$$

Equation 4.70 is the one to use when forward surviving age groups wider than a single year. An important application of Eqs. 4.69 and 4.70 for forward survival of age groups in year y to corresponding age groups in year $y+n$ is the ***projection of populations into the future***. The crux of projecting a population is to take that population at a given date (normally a census date) and survive it to some future date n years beyond the first date by applying the appropriate forward survival equation to each of its constituent age groups. Having survived our census population in year y to year $y+n$, there are two other general elements to be attended to in order to complete our projection.

1. We must consider ***what is likely to happen to births*** over the projection period. Clearly the youngest members of our census population in year y (those aged 0 years) will be aged n years at the end of the n-year projection period. Younger members of our projected population will be ***survivors of birth cohorts born between the census and the projection date***. We must therefore estimate the sizes of these birth cohorts and then survive ***those*** to the projection date to get the projected population aged 0 to $n-1$ years. Of course most of the births that will take place over the projection period ***have yet to take place*** at the time the projection is made. It follows that an integral part of the projection exercise is making assumptions about the sizes of birth cohorts (i.e., about fertility trends) over the projection period. Commonly a ***series*** of projections is made, each incorporating a different assumption about likely fertility trends, so that a range of options as to what the future population size and distribution by age might be is established.

   The type of exercise just described entails ***forward survival from birth***. If this is done for 1-year birth cohorts (which will form 1-year age groups in the projected population) we equate the righthand sides of Eqs. 4.63 and 4.67:

   $$P(y)_x/B\,(y-x-1) = L_x/l_0$$

   and rearrange to get:

   $$P(y)_x = B\,(y-x-1)\,.\,(L_x/l_0) \tag{4.71}$$

   Where y is the projection year; x is a single-year age group in the projected population such that $x < n$ (where n is the projection period in years); $B(y-x-1)$ is estimated births in the year commenced the date in year y to which the population projection pertains (e.g., mid-year) in year $y - x - 1$; L_x comes from an appropriate single-year-of-age life table.

If forward surviving t-year birth cohorts (which will form t-year age groups in the projected population) we equate the righthand sides of Eqs. 4.64 and 4.68:

$$P(y)_{x \text{ to } x+t-1}/B^{t,y-x-t} = {}_tL_x/t \,.\, l_0$$

and rearrange to get:

$$P(y)_{x \text{ to } x+t-1} = B^{t,y-x-t} .\, ({}_tL_x/t \,.\, l_0) \tag{4.72}$$

Where y is the projection year; x is the lower limit and t is the width of an age group in the projected population to be estimated using forward survival from birth (it is necessary that $x + t - 1 < n$, where n is the projection period in years); $B^{t,y-x-t}$ is estimated births during a t-year period commenced the date in year y to which the population projection pertains (e.g., mid-year) in year $y - x - t$; ${}_tL_x$ comes from an appropriate abridged life table.

Equations 4.71 and 4.72 can be used to forward survive birth cohorts born during the n-year projection period which will form ages 0 to $n - 1$ in the projected population. Figure 4.8 illustrates how population projection involves both forward survival of a census population and forward survival of projection period birth cohorts.

2. We must consider whether, and if so how, ***migration*** is likely to modify our population over the projection period. In other words, in addition to making assumptions about future fertility we must make assumptions about future migration. These assumptions can be important when projecting populations at the ***national*** level, but are likely to be even more critical when making projections at the ***sub***-national level (i.e., for regions of a country, cities, local government areas, etc.).

   As with making assumptions about future fertility, it is common in population projection exercises to prepare ***series*** of projections based on a range of migration assumptions. The mathematics of incorporating migration into population projections are not dealt with here. It is sufficient at this juncture to appreciate the need in projecting populations to weld onto the forward survival of an initial census population and births during the projection period a consideration of likely migration trends and patterns during that projection period. Population projection is dealt with in greater detail in Chap. 9.

Estimation of Births

Just as we can use life table survival ratios to project populations into the ***future*** using ***forward*** survival, so we can use them to work ***backwards*** from a population

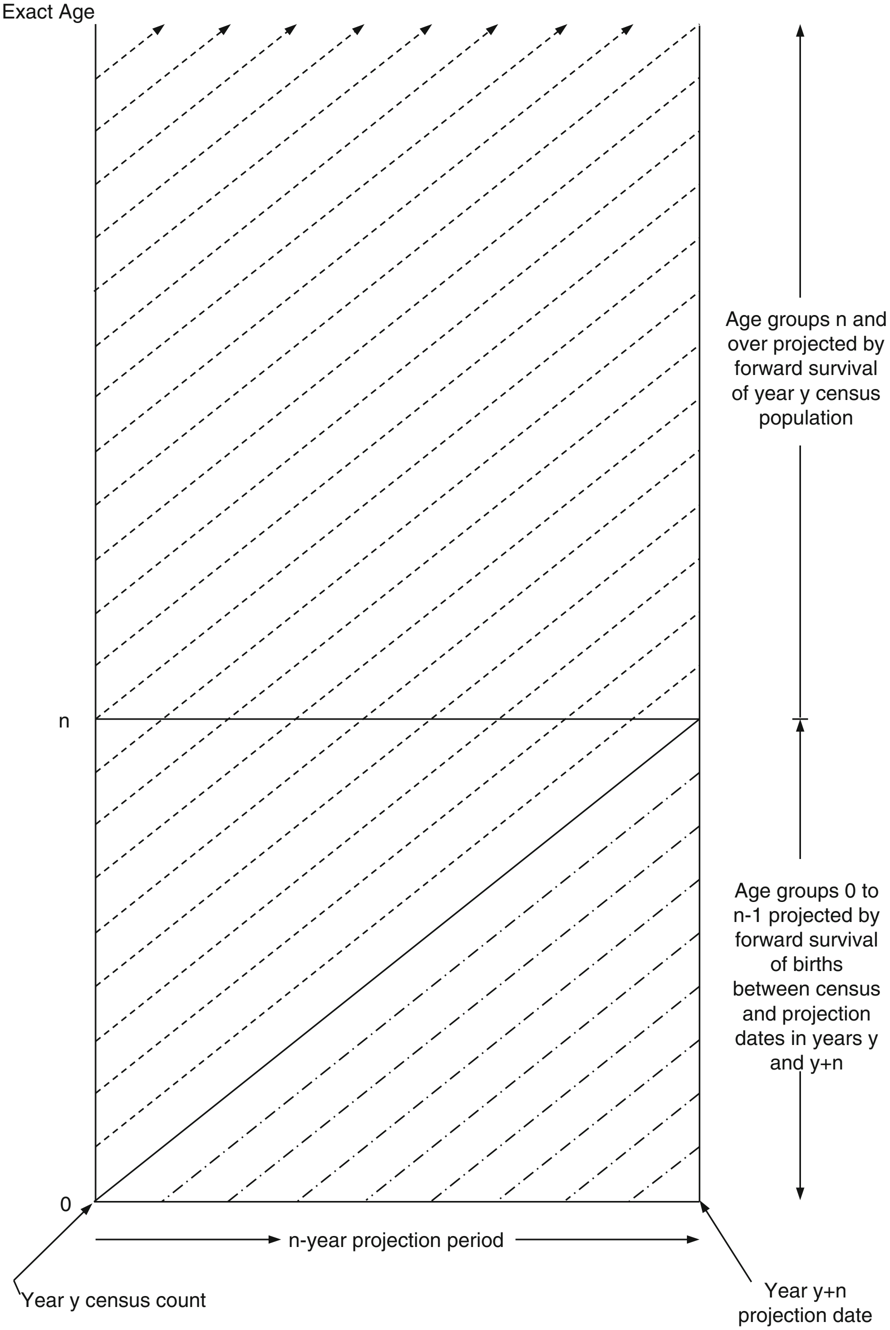


Fig. 4.8 Schematic Lexis representation of the role of forward survival in population projection

distribution by age to obtain information about the past. This procedure is known as ***REVERSE SURVIVAL***.

We could, if we wished, use this method to derive a population distribution by age at some ***earlier*** point in time, assuming no migration and with the proviso that only age groups for which there were survivors in the later population would be covered. We might, for example, be interested in estimating the size and age distribution of a population at some date ***between*** two censuses. We could use a combination of forward survival from the earlier census and reverse survival from the later one, averaging the results of the two exercises. Or if the later census was felt to be more reliable we might rely solely on reverse survival. We could even use reverse survival from that census to assess the extent of unreliability in the earlier, less reliable, census.

A more useful application of reverse survival, however, is in ***estimating births for the period preceding the date of a population count*** (normally a census). This technique uses the idea of survival from birth in reverse, and is especially useful when dealing with populations for which there are no birth registration data or for which birth registration data are incomplete. It enables us to estimate births from census data. However:

1. The technique is only as good as the quality of the census and the life table one uses.
2. If birth registration is incomplete or non-existent, so may death registration be. This may impede the generation of a life table for the population of interest, although this problem can be overcome by resorting to an appropriate ***model*** life table. Obviously, when this approach is taken, the accuracy of birth estimates will depend on how well the model table summarizes mortality conditions in the population of interest.

Suppose a census is conducted in year y. Then under an assumption of zero migration we have, equating the righthand sides of Eqs. 4.63 and 4.67:

$$P(y)_x / B\,(y - x - 1) = L_x / l_0$$

This is the same equation as the one from which we developed Eq. 4.71 for forward survival of a birth cohort to age x. We again rearrange it, but this time treating B(y – x – 1) as the unknown and making it the subject of the new equation:

$$B\,(y - x - 1) = P(y)_x . \,(l_0 / L_x) \tag{4.73}$$

Where y is the census year; x is a single-year age group at the census; $B(y - x - 1)$ = births over the 1-year period commenced the date of the census count in year $y - x - 1$; L_x comes from an appropriate single-year-of-age life table.

So, for example, from the population aged 5 at a census held on 1 April 2006 we can estimate the number of births during the year commenced 1 April 2000 ($y = 2006$, $x = 5$ so that $y - x - 1 = 2000$).

$$B(2000) = P(2006)_5 . (l_0/L_5)$$

Note that, except where the population (census) count was taken at midnight on 31 December, estimates of births obtained from this procedure ***do not relate to calendar years***. They relate to years commenced the date (day and month) when the population count was taken. However, having estimated births for a series of successive non-calendar years it is possible to convert these to estimates for calendar years assuming even distributions of births through non-calendar years.

In Eq. 4.73 we have a method for estimating births over a 1-year period through reverse survival of a 1-year age group. It is equally possible to estimate births over a t-year period through reverse survival of a t-year age group. Suppose again that a census is conducted in year y. Under an assumption of zero migration we have, equating the righthand sides of Eqs. 4.64 and 4.68:

$$P(y)_{x \text{ to } x+t-1}/B^{t,y-x-t} = {}_tL_x/t \,.\, l_0$$

This is the same equation as the one from which we developed Eq. 4.72 for forward survival of a t-year birth cohort to age group x to $x + t - 1$. We again rearrange it, but this time treating $B^{t,y-x-t}$ as the unknown and making it the subject of the new equation:

$$B^{t,y-x-t} = P(y)_{x \text{ to } x+t-1} . (t \,.\, l_0/{}_tL_x) \tag{4.74}$$

Where y is the census year; x is the lower limit and t is the width of an age group at the census; $B^{t,y-x-t}$ = births over the t-year period commenced the date of the census count in year $y - x - t$; ${}_tL_x$ comes from an appropriate abridged life table.

So, for example, from the population aged 0–4 at a census held on 1 April 2006 we can estimate the number of births during the 5-year period commenced 1 April 2001 ($y = 2006$, $x = 0$, $t = 5$ so that $y - x - t = 2001$).

$$B^{5,2001} = P(2006)_{0-4} . (5 \,.\, l_0/{}_5L_0) = P(2006)_{0-4} . [5 \,.\, l_0/\, ({}_1L_0 + {}_4L_1)]$$

Since ${}_5L_0 = {}_1L_0 + {}_4L_1$, because an abridged life table begins with a 1-year, then a 4-year age interval.

Equations 4.73 and 4.74 for estimating numbers of births by reverse survival from a census have assumed zero net migration. In using reverse survival for this purpose it is important to assess how reasonable the zero migration assumption is. Normally there isn't much of a problem when estimating births over the few years immediately preceding a census. This is done by reverse survival of the ***youngest*** age groups at the census, whose youth means that there has been limited opportunity for migration, although obviously parental migration can draw in children as well.

However, more importantly, as one attempts to estimate births further back in time one reverse survives older and older age groups at the census, which have had more and more time in which to be influenced by migration. The zero migration assumption becomes increasingly problematic, and one's estimates of births become increasingly prone to error. The technique is thus best reserved for estimating births in the few years immediately preceding a census, unless a credible basis exists for incorporating migration into the calculation.

Estimation of Intercensal Migration

The estimation of intercensal migration is another important application of life table survival ratios. The general strategy is to forward survive an age group from an earlier census to find the ***expected*** size of the corresponding age group at a later census ***under an assumption of zero migration***. This is then compared with that age group's ***actual*** size at the later census, the difference between the two being net migration.

Using forward survival we have, for single-year age groups:

$$P^*(y)_x / P(y-n)_{x-n} = L_x / L_{x-n}$$

Where $P(y-n)_{x-n}$ is the population aged $x-n$ at a census in year $y-n$ which marks the ***beginning*** of an n-year intercensal period; $P^*(y)_x$ is the ***expected*** number of survivors from $P(y-n)_{x-n}$ at the end of the intercensal period assuming no migration; L_x and L_{x-n} come from a single-year-of-age life table which summarizes intercensal mortality conditions. It might also help to realize that $y =$ the later census year, $x =$ the relevant single-year age group at the later census, and $n =$ the width of the intercensal period in years.

It follows that:

$$P^*(y)_x = P(y-n)_{x-n} . (L_x / L_{x-n}) \qquad (4.75)$$

Whence net migration over the intercensal period for the cohort aged x at the later census is given by:

$$NM = P(y)_x - P^*(y)_x \qquad (4.76)$$

This is the actual, or ***observed***, population aged x at the later census less the population aged x ***expected*** had there been no intercensal migration.

If our interest is in estimating intercensal migration for t-year age groups (rather than single-year age groups) we have, using forward survival:

$$P^*(y)_{x \text{ to } x+t-1} / P(y-n)_{x-n \text{ to } x+t-1-n} = {}_tL_x / {}_tL_{x-n}$$

Where $P(y-n)_{x-n \text{ to } x+t-1-n}$ is the population aged $x-n$ to $x+t-1-n$ at a census in year $y-n$ which marks the beginning of an n-year intercensal period; $P^*(y)_{x \text{ to } x+t-1}$ is the expected number of survivors from $P(y-n)_{x-n \text{ to } x+t-1-n}$ at the end of the intercensal period assuming no migration; ${}_tL_x$ and ${}_tL_{x-n}$ come from an abridged life table which summarizes intercensal mortality conditions over t-year age intervals (they can also be obtained by summing values of L_x from an appropriate single-year-of-age life table over relevant t-year age intervals). It might also help to realize that $y =$ the later census year, $x =$ the lower limit of the relevant t-year age group at the later census, and $n =$ the width of the intercensal period in years.

It follows that:

$$P^*(y)_{x \text{ to } x+t-1} = P(y-n)_{x-n \text{ to } x+t-1-n} \cdot ({}_tL_x / {}_tL_{x-n}) \tag{4.77}$$

Whence net migration over the intercensal period for the cohort aged x to $x+t-1$ at the later census is given by:

$$NM = P(y)_{x \text{ to } x+t-1} - P^*(y)_{x \text{ to } x+t-1} \tag{4.78}$$

This is the actual, or ***observed***, population aged x to $x+t-1$ at the later census less the population aged x to $x+t-1$ ***expected*** had there been no intercensal migration.

As an example, let us assume we have censuses conducted on 1 April 1996 and 1 April 2006, and that we want to specify equations following Eqs. 4.77 and 4.78 which will permit net intercensal migration for the cohort aged 20–24 at the 2006 census to be estimated. We have that $y = 2006$ (the later census year), $x = 20$ (the lower limit of the relevant age group at the later census), $t = 5$ (the width of the relevant age group) and $n = 10$ (the length of the intercensal period). Therefore, following Eq. 4.77:

$$P^*(2006)_{20-24} = P(1996)_{10-14} \cdot ({}_5L_{20} / {}_5L_{10})$$

And, from Eq. 4.78:

$$NM = P(2006)_{20-24} - P^*(2006)_{20-24}$$

Note that a rather intimidating equation like Eq. 4.77 looks a good deal less intimidating once we get down to a real example. The expected population aged 20–24 in 2006 equals the observed population aged 10–14 (10 years younger) in 1996 (10 years earlier) multiplied by a life table survival ratio which is the ratio of the sizes of the same two age groups in the life table stationary population (the ${}_nL_x$ values for those age groups).

Equations 4.75, 4.76, 4.77 and 4.78 facilitate the calculation of ***intercensal*** net migration as a cohort ages from a younger age ***group*** at the earlier census to a corresponding older age ***group*** at the later census. It is in theory also possible to estimate net migration as a cohort ages between exact ages, between an exact age and an age group, or between an age group and an exact age. In practical demographic terms the problem of this nature most likely to be of interest is one requiring net migration between ***birth*** (exact age 0) and the relevant age group at a later census to be estimated for a cohort. The principle is the same as before. We obtain the ***expected*** size of the age group at the census, but this time using the principles of forward survival ***from birth***, and compare this with the ***observed*** size of the age group.

For net migration from birth to a single-year age group x in year y we have, equating the righthand sides of Eqs. 4.63 and 4.67 with $P^*(y)_x$ written for $P(y)_x$:

$$P^*(y)_x / B(y - x - 1) = L_x / l_0$$

Where $B(y-x-1)$ is births during the year commenced the date of the year y census count in year $y-x-1$; $P^*(y)_x$ is the ***expected*** number of survivors from birth forming age group x at the census in year y assuming no migration.

Whence:

$$P^*(y)_x = B(y - x - 1) \,.\, (L_x / l_0) \tag{4.79}$$

After which net migration would be given by Eq. 4.76.

And for net migration from birth to a t-year age group x to $x + t - 1$ in year y we have, equating the righthand sides of Eqs. 4.64 and 4.68 with $P^*(y)_{x \text{ to } x+t-1}$ written for $P(y)_{x \text{ to } x+t-1}$:

$$P^*(y)_{x \text{ to } x+t-1} / B^{t,y-x-t} = {}_tL_x / t \,.\, l_0$$

Where $B^{t,y-x-t}$ is births over the t-year period commenced the date of the year y census count in year $y-x-t$; $P^*(y)_{x \text{ to } x+t-1}$ is the ***expected*** number of survivors from birth forming age group x to $x+t-1$ at the census in year y assuming no migration; x is the lower limit and t is the width of this age group at the census.

Whence:

$$P^*(y)_{x \text{ to } x+t-1} = B^{t,y-x-t} \,.\, ({}_tL_x / t \,.\, l_0) \tag{4.80}$$

After which net migration would be given by Eq. 4.78.

Unless the year y census to which net migration from birth was being estimated happened to have taken place at midnight on 31 December, the birth cohorts corresponding to age cohorts at the census for which migration estimates were sought would be ***non-calendar year birth cohorts***. Unless detailed data on births by month/week/day were available, the sizes of these cohorts would need to be estimated from data on births during calendar years by assuming even distributions of births by date of occurrence through those years.

It is perhaps worth concluding this section by recalling that life tables usually are constructed separately for males and females. Thus in using life tables to estimate intercensal net migration, to estimate births or to project a population we normally have to deal separately with the male and female components of the population.

Survival Between Birthdays (Exact Ages): Joint Survival

We have looked at survival from one age ***group*** to another, and at survival from ***birth*** (exact age 0) to an age ***group***. In the former instance a life table survival ratio was the ratio of two L_x values; in the latter it was the ratio of an L_x value to l_0. It is also possible to calculate survival ratios which measure the probability of surviving between two ***birthdays***, or exact ages, and consistent with the principle enunciated earlier, such survival ratios are the ratio of two life table l_x values. The probability of surviving from one's xth birthday to one's x + nth birthday (from exact age x to exact age x + n) is given by:

$$ {}_{n}SR_{x} = l_{x+n}/l_{x} \tag{4.81}$$

This application of the life table tends to be of actuarial significance more than of demographic significance, but it does raise the issue of ***joint survival***. Joint survival refers to the probability that two (or more) persons will both (all) survive over a given period, or indeed that they will survive and ***fail*** to survive in some nominated combination. ***It is measured by the product of the probabilities of survival or failure to survive for each individual person***.

As an example, suppose a man is now aged 30 years 3 months and his wife is aged 27 years 9 months. What is the probability that the husband will survive to retirement age (his 65th birthday) with his wife still alive? The probability that the husband survives a further 65–30.25 = 34.75 years is given by:

$$ {}_{34.75}SR_{30.25} = l_{65}/l_{30.25} $$

Where l_{65} and $l_{30.25}$ are taken from an appropriate life table for ***males***.

The probability that the wife survives over the same period of 34.75 years, at the end of which she will still be 2 years 6 months younger than her husband is given by:

$$ {}_{34.75}SR_{27.75} = l_{62.5}/l_{27.75} $$

Where $l_{62.5}$ and $l_{27.75}$ are taken from an appropriate life table for ***females***.

To find the probability of joint survival to the husband's retirement we would evaluate these two survival ratios and multiply them together. The spectre of l_x values where x is not a whole number will be new. They can be evaluated by linear interpolation between the l_x values of a single-year-of-age life table, since other than at very young ages it is reasonable to assume linear survivorship between exact age x and exact age x + 1. So, for example, $l_{30.25}$ is one-quarter of the way between l_{30} and l_{31}.

If you are having difficulty imagining 'demographic' applications of this type of calculation, think in terms of studies of phenomena such as widowhood, orphanhood, the likelihood as a parent of experiencing the death of a child, and the effects of mortality on family composition (e.g., in producing sole parent families). These are all topics that can be greatly illuminated by illustrative joint survival calculations, especially comparatively over time or between different populations at a point in time.

Further Issues in the Analysis of Mortality

In the final section of this chapter we look briefly at several further issues that become relevant as mortality analysis is expanded beyond the basics covered thus far. The aim is simply to alert you to these issues, not to develop them in detail. The issues in question are: demographers' interest in causes of death when undertaking detailed studies of mortality trends and patterns; the nature of differential mortality analysis; the concept of perinatal mortality; more sophisticated types of life tables; the concept of years of potential life lost; and approaches to studying not just the quantity of life lived (survivorship), but the quality of life lived (i.e., taking morbidity as well as mortality into account in assessing the health of populations).

Causes of Death

It is a feature of detailed studies of mortality trends and patterns that they often focus on different causes of death. In theory, at least, such a focus has the potential to help explain changes in mortality levels apparent from more general studies of life tables. One can sometimes identify increases or decreases in mortality from particular causes as the major contributors to general mortality trends, thereby enhancing understanding of those trends. It may even happen that a cause-of-death study may show rather more change in the pattern of mortality than an examination of all-cause life tables would lead one to suspect, with offsetting changes having occurred in different cause categories. Cause-specific studies of mortality, in which the focus is exclusively on deaths from a particular cause or group of causes, are also possible. The obvious example in recent decades has been studies of HIV-AIDS, which has wreaked havoc with the mortality levels and patterns, and hence with the life tables, of many developing countries since the 1980s.

Classifications of death by cause are normally based on an international classification standard established by the World Health Organization (WHO). The WHO's

International Classification of Diseases, Injuries and Causes of Death (ICD) is revised every 10 years, a reality that can create complications for trend analysis of cause of death data. Deaths that over one period were allocated to one category suddenly, because the opinion of medical personnel with input into the classification system has changed, are allocated to another, and it is not always possible to reclassify data on a consistent basis for analysis purposes. Thus, always approach time series data on causes of death, especially detailed time series data, with caution, and check the consistency of the basis of classification over time. Be especially suspicious of sudden change in the level of mortality from a particular cause; the first question to ask yourself in such a circumstance is, ‘Is this change likely to be an artefact of a change in classification procedures?’ Classifying deaths by cause can be problematic at the best of times, but when deaths that once went into one category suddenly are switched to another life becomes doubly difficult.

Many countries, including most African countries and some Asian countries, do not maintain systems of death certification requiring a medical practitioner to fill out a ‘Certificate of Cause of Death’. For such countries comprehensive analyses of mortality by cause of death are obviously impossible. Some data may be available from records of individual medical establishments, but the potential for bias in these due to factors such as their pertaining to geographically local areas and perhaps only to persons who die in a hospital is acute.

A second problem with cause of death data is that the quality of classification by cause of death can be extremely variable, both from country to country and over time. Data on causes of death published in the 2011 edition of the *United Nations Demographic Yearbook* are illustrative. The fairly broad ICD categories used include one labelled ‘Symptoms, signs and abnormal clinical laboratory findings not elsewhere classified’. Variable proportions of deaths allocated to this essentially ‘dump’ category are reflective of the variable quality of cause of death certification in different countries. In Australia in 2006, for example, just 1.2 % of male deaths and 1.0 % of female deaths were allocated to this category. In South Africa in 2009 the equivalent figures were respectively 12.6 % and 14.8 %, in El Salvador in 2008 they were 14.2 % and 19.5 %, in Sri Lanka in 2006 they were 13.4 % and 21.0 %, and in Thailand in 2006 they were fully 33.5 % and 44.7 %. These differences are clear evidence of considerable disparity in the capacities and/or willingness of the medical systems in different countries to accurately certify causes of death. Problems can especially arise with deaths in old age, which are apt to be attributed to medically imprecise ‘causes’ like ‘senility’. Indeed this may explain why, in the pairs of figures just quoted (except that for Australia) the female figure is always higher. Female mortality, because women are less prone to premature death through accidents, violence and conditions that reflect unhealthy lifestyles (smoking, alcohol consumption, etc.), tends to be more concentrated at advanced ages where deaths are expected, and being precise as to their causes seems consequently in some jurisdictions to not be a high priority.

Do not imagine, however, that even data for a country like Australia are problem-free. A person may suffer from several medically adverse conditions in the period immediately prior to death. In such circumstances the medical practitioner certifying the death, while perhaps listing a ***range*** of causes, is normally required to nominate from among them the ***underlying*** cause of death, which is the one taken notice of in classifying deaths by cause. It should be recognised that this choice may be a matter of medical opinion or even guesswork. Then again, a doctor may describe in intricate medical detail the injuries sustained by an accident victim or a victim of violence, without saying anything about the cause of the accident or the nature of the violence which occasioned death. A statistical agency may have other information, such as a coroner's report, to assist in detecting and correctly classifying such cases, but otherwise misclassification can easily occur. Suicide can be an especially difficult cause of death to detect – if someone deliberately has a fatal motor vehicle accident, for example, and fails to leave any indication of premeditation, their death will probably be deemed to be due to a motor vehicle accident rather than to suicide. We can also conceive of the notion of the 'underlying cause' of a death operating at different levels. At one level the cause may be long-term exposure to an environmental hazard – radioactivity or the dust encountered in working as a miner, for example – or the long-term practice of a dangerous personal habit – smoking or excessive drinking, perhaps. At another level the cause may be poverty, which precluded access to an unpolluted water supply, to adequate knowledge of preventive health behaviour, and to basic medical facilities. At yet another level the cause may be cancer, cholera or coronary disease, or at still another it may be that the heart stopped beating and the brain ceased its activity. Thus, ***even in the best of statistical environments, cause-of-death statistics have a certain fuzziness about them***, and it is important to deal with them warily because of that.

In making comparisons of mortality levels by cause of death it is always important to be aware of 'differences' that are wholly, or substantially, due to differences in age structure. ***Cause-specific death rates should routinely be standardized for age***, for just as general mortality levels are highly dependent on age structure, so, too, are rates of death from particular causes. In a comparison of cause of death data for Sri Lanka in 1986 and Australia in 1994, for example, the female death rate from 'Other diseases of the digestive system' was 11.0 per 100,000 population in Australia and 5.3 per 100,000 in Sri Lanka. Age-specific rates were, however, except at ages 75 and over, all higher for Sri Lanka. The exception at the oldest ages was probably due partly to a tendency to attribute deaths at those ages to 'senility' in Sri Lanka, and partly to an older age structure ***within*** the '75 and over' age category for Australia. But even setting these considerations aside, standardization of the Australian death rate to the age structure of the Sri Lankan population reduces it to 3.5 per 100,000, well below, rather than more than double, the Sri Lankan rate. You will often, in cause of death and other mortality studies, encounter the ***standardized mortality ratio***, a measure introduced in Chap. 2 as a particular example of standardized ratios, measures that facilitate legitimate comparison of summary measures after controlling for differences in population composition. The standardized mortality ratio for mortality from 'Other diseases of the digestive system' among Australian and Sri Lankan females in the present example was 3.5/5.3 = 0.66; i.e., the level of mortality from this cause in Australia in 1994 was two-thirds of what it was in Sri Lanka in 1986. The unstandardized mortality ratio was $11.0/5.3 = 2.08$, suggesting falsely that female mortality from this cause in Australia was more than double what it was in Sri Lanka.

Analysis of Differential Mortality

Studies of differential mortality establish and attempt to explain different levels of mortality in discrete subgroups of a population. This notion really gets back to compositional variables other than age and sex by which it was noted earlier that mortality levels often vary. Thus, studies have established differential risks of dying by ethnicity, geographic location (reflecting environmental factors), occupation (reflecting risk factors associated with particular types of employment), marital status (marriage has been claimed to select the healthy), education (the better educated supposedly are better informed in matters of preventive health care and more knowledgeable of how to seek medical help when needed), parental educational attainment (which can affect infant and child mortality in similar ways), and income (which affects ability to access medical services). In high income countries mortality differentials may also be associated with what have been termed 'social roles', or particular lifestyles. These may reduce or increase risks of death from conditions related to, for example, excessive alcohol consumption, smoking, the taking of other drugs and diet, and from involvement in accidents and violence of various descriptions.

Studies of differential mortality are undertaken largely using techniques of standardization and life table measures, so you are already quite well equipped to tackle them. Some do, however, make use of more sophisticated life table techniques – for example, multiple increment-decrement life tables which, in relation to a variable like marital status, permit changes of marital status over the life cycle to be dealt with in conjunction with cross-marital status differences in mortality risks. These are dealt with briefly below.

Perinatal Mortality

At the beginning of this chapter we defined death as the cessation of life in a person ***born alive***. Demographers also have an interest in ***foetal*** deaths. More commonly referred to as ***stillbirths***, these are defined as involving ***children born weighing at least 500 grams (or if body weight is unavailable, at least 22 weeks gestation) who did not, at any time after being born, breathe or show any other sign of life***. The arbitrary limits of 500 g or 22 weeks gestation are supposed to signify ***theoretical viability*** – a theoretical ability to survive outside the uterus.

One reason demographers take an interest in stillbirths is that whether or not a child breathes after birth is often a fine line; there frequently isn't much to distinguish a stillbirth from an early infant death, and it can be debatable which category a particular child should be classified to. Treating the two groups together conveniently avoids this classification problem. Secondly, stillbirths and very early infant deaths often have similar causes, and thus are likely to respond similarly to antenatal medical initiatives, be they ones that enhance survival chances or ones (such as the prescribing of thalidomide to pregnant women to alleviate morning sickness in the 1950s and early 1960s) that have the opposite effect (thalidomide was later shown to cause severe birth defects). On both of these counts, too, their simultaneous treatment is recommended.

We have noted several times the tendency for infant deaths to concentrate very early in the first year of life. Those that occur within 28 days following birth are given a special name. They are termed ***neonatal*** deaths, and within this category those that occur within 7 days of birth are termed ***early neonatal*** deaths. In the broadest definition, it is to the amalgam of foetal deaths (stillbirths) and neonatal deaths that the term ***perinatal mortality*** is given. By this definition a perinatal death is ***a foetal death (stillbirth) or a neonatal death***. One does, however, encounter variations in the definition of a perinatal death. Sometimes, for example, only ***early*** neonatal deaths are included, while in the other direction only ***late*** foetal deaths (those occurring after the 28th week of gestation) may be defined as perinatal. Two things follow. First, ***in general terms*** perinatal mortality is concerned with deaths of children occurring just before, during, and just after childbirth (or ***parturition***). Second, one must always be careful to check the specific definition underlying perinatal mortality data, especially if comparative analysis between populations or over time is contemplated.

Perinatal death rates are ratios of perinatal deaths to ***total*** births (i.e., live births plus stillbirths). They often are divided into stillbirth (or late foetal death) rates and neonatal (or early neonatal) death rates, but the latter can be a source of confusion. Just as, in the broadest definition, perinatal deaths can be divided into ***foetal*** deaths and neonatal deaths, so, too, infant deaths can be divided into neonatal deaths and ***post***-neonatal deaths. When neonatal (or early neonatal) death rates are calculated in the contexts of perinatal and infant mortality ***their denominators are different***, respectively ***in***cluding and ***ex***cluding stillbirths (or late foetal deaths). Thus, you should always check which denominator was used to calculate such death rates. It is more usual for them to be calculated as a component of infant mortality, and hence relative to live births only, but this is not invariably the case. Be careful, in particular, not to compare neonatal or early neonatal death rates when one figure has a live births denominator and the other has a total births denominator.

More Sophisticated Life Tables

We have dealt in this chapter with standard, or single-year-of-age, life tables and abridged life tables as vehicles for the study of mortality, or survival. These are examples of what are known as ***single decrement*** life tables, which deal with an

attrition process (recall Chap. 3) in which only one transition is of interest – from the ***transient*** state 'alive' to the ***absorbing*** state 'dead'. The state 'alive' is said to be 'transient' because nobody remains in it indefinitely. The state 'dead' is by contrast 'absorbing' because, once entered, there can be no transition back in the reverse direction (at least not for mere mortals).

Without going into detail it is appropriate at this juncture to alert you to two things, one of which was touched on in passing in Chap. 3. First, ***life tables can be used to study processes other than mortality***. Thus when we discussed attrition in Chap. 3 the example we used focused on breastfeeding mothers transitioning from being breastfeeders to having terminated breastfeeding. We could similarly use life table techniques to study the process of transition with age from being never married to being married, or the process of transition with marriage duration from being married to being divorced. Second, ***life tables can be used to study processes where transitions to more than one absorbing state are possible, and also to study processes allowing transitions to and fro between multiple transient states as well as to one or more absorbing state***. Life tables that permit the former are known as ***multiple decrement*** life tables. Those that permit the latter are ***multistate*** life tables.

Multiple Decrement Life Tables

The notion of multiple decrement life tables really comes to mind once one appreciates that life tables can be used to study processes other than mortality. Clearly, for example, if we are thinking about attrition from a radix of never married males or females through marriage, there is the possibility, indeed the near certainty, that some will be lost along the way to death prior to marriage. Hence we are thinking in terms of a ***double decrement***, and in the case of marriage about what is known as a ***net nuptiality table*** – one that captures the attrition of a never married radix population through either marriage ***or*** death as a single person. Sooner or later everyone ceases to be a member of the never married population, and they do this either by marrying or by dying. Marriage is an absorbing state in this context because, once married, one can never revert to being 'never married'.

A common application of ***multiple decrement*** life tables (three or more modes of decrement) is to the study of mortality patterns by cause of death. Instead of simply studying the transition from 'alive' to 'dead' we can study transitions from 'alive' to 'dead from cause A', 'dead from cause B', 'dead from cause C', etc. These alternative absorbing states are known as ***competing risks***.

Construction of a multiple decrement life table that recognizes different causes of death, or groups of causes of death, ***begins with an all-causes standard or abridged life table for the relevant population***. The values of ${}_nd_x$ in this life table are then partitioned into values ${}_nd_{x,1}$, ${}_nd_{x,2}$, ... ${}_nd_{x,c}$ (where c is the number of discrete cause of death categories to be recognized in the analysis) on the basis of the proportions of total observed deaths at each age attributed to each cause in the relevant population in the year (or group of consecutive years) to which the all-causes life table pertains.

Thus:

$$ {}_{n}d_{x,c} = {}_{n}d_{x}\,({}_{n}D_{x,c}/{}_{n}D_{x}) \tag{4.82}$$

Where D stands for observed deaths in the relevant population.

These columns of life table deaths partitioned by cause are then used to construct life tables conditional on dying from each cause. The radices (plural of 'radix') of these conditional life tables are given by the sums of the relevant columns of life table deaths, which means that they sum to 100,000, the radix of the initial all-cause life table (i.e., l_0 is partitioned into additive components $l_{0,1}, l_{0,2}, \ldots l_{0,c}$). Other values in these columns are then found from:

$$ l_{x+n,c} = l_{x,c} - {}_{n}d_{x,c} \tag{4.83}$$

with columns so generated describing the survival experience of persons who eventually die from cause of death c.

Cause-specific probabilities of dying are given by:

$$ {}_{n}q_{x,c} = {}_{n}d_{x,c}/l_{x} \tag{4.84}$$

Note that the denominator on the righthand side of this equation is l_x (from the initial all-causes life table), ***not*** $l_{x,c}$ (from the cause-specific life table for cause c). It follows, since the sum of values of ${}_{n}d_{x,c}$ across all cause of death categories c is ${}_{n}d_{x}$ from the initial all-causes life table, that the sum of values of ${}_{n}q_{x,c}$ across all cause of death categories c is ${}_{n}q_{x}$ from the initial all-causes life table. That is, for a given age interval the probabilities of dying by cause sum to the all-causes probability of dying.

Other columns of the cause-specific life tables (${}_{n}L_{x,c}$, $T_{x,c}$ and $e^{o}_{x,c}$) are generated using the same equations as are used to generate equivalent all-cause life table functions, but with all terms being specific for cause c. Columns of $e^{o}_{x,c}$ values give expectations of life remaining ***for persons who will ultimately die of cause c***.

Multistate Life Tables

Multistate life tables are mentioned only in passing. Also known as ***increment-decrement life tables***, they are a type of life table you may encounter reference to in the literature, so it is as well to have a basic understanding of what they are. Their distinguishing feature is that they have two or more transient states, for each of which transitions ***both to and from*** other transient states (although not necessarily all of them) are possible. They may also feature one or more absorbing states, but the label 'increment-decrement life tables' captures their distinctive characteristic – their allowing for flows of population both into as well as out of some states, instead of in only one direction.

Multistate life tables are not used for studies of mortality *per se*, because death is the ultimate absorbing state. They can be used to study health status, with population moving between states defined by being diagnosed with, or declared recovered from, various health conditions, as well as, potentially, to the absorbing state deceased. More commonly, however, they are used in analyses of (i) marital statuses and relationships (where individuals, over time, can move into and out of states such as 'married', 'divorced', 'widowed', 'in a relationship' and 'not in a relationship' multiple times), (ii) internal migration (where multiple movements into and out of geographic regions are possible), and (iii) labour force participation (where people can move into and out of the labour force, and back and forth between full-time and part-time employment, multiple times over their working lives). Multistate life tables designed to study labour force activity are often known as ***working-life tables***.

Competing Risks and Cause-Eliminated Life Tables

The decrement probabilities in multiple decrement life tables are described variously as ***crude***, or ***dependent*** probabilities. Their values for a particular mode of decrement, or risk, depend partly on those for the competing risks. If probabilities of succumbing to one risk change over time or differ between two populations, that has implications for probabilities of succumbing to other risks. Where the competing risks are alternative causes of death, for example, because everyone must die from one cause or another, rising or higher dependent probabilities of dying from one cause or group of causes must be compensated by falling or lower dependent probabilities of dying from one or more other causes or groups of causes, and vice versa.

Often the real interest is in the ***independent*** or ***underlying*** probabilities associated with a particular risk; that is, in the probabilities that would apply ***in the absence of competing risks***. In the context of the analysis of mortality by cause of death, such probabilities, and ***cause-eliminated*** life tables constructed using them, provide a sounder basis for comparison of mortality from particular causes in different populations or over time. They also provide a sounder approach to the assessment of past mortality patterns as a basis for projecting cause-specific mortality, and thereby future life expectancy. Independent probabilities of dying from a particular cause are invariably higher than corresponding dependent probabilities, since people die from that cause who, in the presence of competing risks, would die earlier from some other cause.

Cause-eliminated life tables of the type just described eliminate ***all*** causes of death other than the one that is the focus of attention. Other variants of this procedure are also followed. In particular it is common to construct life tables in which ***individual*** causes are eliminated, so that improvements in life expectancy that would result from elimination of that cause, or alternatively the reduction in longevity to which the cause gives rise, can be assessed. The effects of cause-elimination on mortality from ***other*** causes can also be studied, as can the implications of ***partial*** elimination of individual causes (addressing questions of the type 'What would be the gain to life expectancy if mortality from cause x could be reduced by y percent at all ages?'). For many if not most causes, ***reduction*** in mortality will be a more realistic scenario to examine than complete elimination. It is also, of course, possible to conceive of cause-specific mortality ***rising***, and of wanting to assess the implications of this type of scenario, too. The obvious contemporary example has been projecting the impact on longevity of the spread of HIV/AIDS, although many countries in recent years have been bringing epidemics more under control.

The mathematics associated with the procedures just summarized are not straightforward, and will not be canvassed here. Anyone keen to pursue them is referred to Chiang (1968: Chapter 11), Smith (1992: 159–169) and Siegel and Swanson (2004: 329–331).

Years of Potential Life Lost (YPLL)

YPLL is a measure that is sometimes calculated for particular causes of death or groups of causes to capture the ***aggregate*** loss of life due to a cause or group of causes. It originated with the argument of Dempsey (1947) that measuring ***time*** lost due to mortality provided more insight into the burden it imposed on society than simply relying on crude or age-standardized death rates, which captured only the ***amount*** of death in a population. YPLL recognized that, beyond the mere occurrence of a death, the ***age*** at which it occurred had an important bearing on its societal implications. The method weighted deaths at younger ages more heavily than those at older ages and thus became a favoured metric for, in particular, advocates for a greater share of health resources to address causes of death that took lives at younger ages. It made those causes look more important in the overall scheme of competing healthcare demands than they otherwise would.

The literature contains a number of methodological variants of YPLL measures (Gardner and Sanborn 1990). In its most straightforward form, the number of deaths from a cause (or group of causes) at each single year of age is multiplied by the average expectation of life remaining at the mid-point of that age interval obtained from an appropriate life table. These values are then summed over all single-year age intervals for which deaths are recorded to yield total YPLL. Mid-interval life expectancies typically are not directly available, but can be estimated as averages of those associated with the upper and lower bounds of each interval. If only an abridged life table is available the same methodology can be applied using the age intervals represented in that life table rather than single-year intervals.

This YPLL approach effectively sets what has variously been described as a 'reference' age or an 'upper cut-off' age for each age interval (the age defining how long a person would be expected to live in the absence of premature death) as the interval midpoint plus average expectation of life remaining at that midpoint. Other approaches have set reference or upper cut-off ages that were constant for deaths in all age intervals, variously at (i) the expectation of life at birth for the population in question, (ii) arbitrary values such as 65, 70, 75 or 85 years, and (iii) ages by which a nominated percentage of all deaths (e.g., 90 %) have occurred. In addition there has been debate over whether deaths at age 0 should be excluded (the consensus seems to be that they shouldn't) and some studies have focused on loss of ***working*** life, excluding life lost before some minimum working age (e.g., 15 years) as well as that lost beyond some upper age limit for working life (e.g., 65 or 70).

These methodological variants need to be appreciated, as do some issues to which they give rise. First, different methodologies are apt to yield different results, and it is important to be mindful of the particular methodology used. Second, the younger the reference or upper cut-off age used the greater will be the weighting in results toward health conditions (e.g., accidents and suicide/homicide) that are more prominent causes of death at younger ages.

More refined calculations of YPLL make use of life expectancies from cause-eliminated life tables rather than from all-causes life tables. Strictly speaking, for a given cause or group of causes, c, one should use in YPLL calculations life expectancies from a life table in which c is eliminated as a cause. The exercise is endeavouring to determine how many more years would be lived by a population if cause c was not a risk, and elimination of that risk will raise life expectancies for all other risks.

Discounts are sometimes introduced to YPLL calculations, the argument for them being that a year of life gained (or lost) immediately is of greater value (or cost) than one gained (or lost) in, say, 20 years' time. The effects of discounting can be substantial. Discounting at 3 % per year, for example, reduces the YPLL due to a male infant death in Australia from 81 years to 30 years. It is therefore important to be aware of whether, and what, discounts have been applied.

A variant on (total) YPLL is to calculate the cost of different causes of death in loss of ***productive*** life. Some might wish to debate the definition of 'productive', but conventionally productive life is defined as ***working*** life, or life before retirement age (exact age 65). The calculation proceeds exactly as before, except that ***partial*** life expectancies (i.e., expectations of life remaining ***before exact age 65***) are used instead of total life expectancies. These partial life expectancies at any exact age are readily calculated. To obtain the average expectation of life remaining before retirement at exact age x, simply sum, in the relevant life table, the L_x columns between exact ages x and 65, and divide by l_x.

Focusing on the Quality of Life: Health Expectancies

As life expectancies in developed countries have increased to historically unprecedented levels in recent decades, demographers and public health practitioners have become increasingly interested not just in the ***quantity*** of life lived (survivorship), but also in its ***quality***. There is interest in the extent to which total life expectancies represent, on the one hand, ***healthy life***, and on the other, ***health-impaired life***. Rising life expectancies give rise to questions as to the extent to which these increases in longevity represent quality life, or merely extensions of health-impaired life. And more recently ***burden of disease*** concepts have sought to extend assessments of the health of populations beyond life table-based analyses of cause-specific mortality to also take into account non-fatal health conditions.

Attempts to introduce notions of quality of life into assessments of population health date from a paper by Sanders (1964), in which he wrote, after expressing dissatisfaction with existing approaches to measuring community health levels:

> *It is proposed, therefore, that we develop a modified life table method of analysis in measuring comparative adequacy of health services for different population groups. In such an analysis we would not only determine for each age the probability of survival, but also the subsidiary probabilities of those surviving on the basis of their functional effectiveness. This would range from individuals who are completely dependent on others, even for carrying on their daily living activities, to those fully equipped to carry on with no apparent handicaps all the functions characteristic of their age and sex. (Sanders* 1964*:* 1067*)*

The first method for calculating ***health expectancies*** as foreseen by Sanders was proposed by Sullivan (1971). Subsequent methodological developments can be traced in publications such as those by Mathers et al. (1994), Robine et al. (1999), Murray et al. (2002), Robine et al. (2003), Albert et al. (2008), Guillot and Yu (2009) and Bruggink (2011). At the core of the concept of ***health expectancy***, or ***disability-free expectation of life*** as it is sometimes called, is the idea of taking the life table notion of total expectation of life, whether at birth or at some other exact age (e.g., 65), and splitting it into expected years of ***disabled, or unhealthy life*** and expected years of ***disability-free, or healthy life – healthy life years (HLY)***. The former health status category is also often further split by, for example, distinguishing (in ascending order of severity) 'impairment' from 'disability' and 'handicap' (various other labels are also given by different studies to subcategories of unhealthy life – e.g., 'moderate activity limitation' and 'serious activity limitation'). The splitting process uses age-specific rates of disability (and impairment, handicap etc.) from cross-sectional surveys to partition the L_x, and hence the T_x, values in a population's life table. This in turn allows total life expectancy to be partitioned by health status according to:

$$e^o{}_{x,h} = T_{x,h}/l_x \tag{4.85}$$

Where h is a health status category.

A major field of debate in this type of analysis is, of course, how one defines and measures health status concepts such as 'disability-free' or 'healthy' on the one hand, and 'impairment', 'disability', 'handicap' and other subcategories falling under the rubric 'unhealthy' on the other. Does one, for example, allow individuals to self-select into these categories, or does one attempt to impose more objective classificatory criteria? If the latter, what questions should be asked as the basis for classification? It naturally follows from these issues that comparability between studies of health expectancy is a major consideration. Full comparability requires common health status categories defined in a standard way based on a standard survey questionnaire.

As an example of the sorts of findings health expectancy analysis can yield, data for the 27 countries of the European Union in 2010 showed that while male life expectancy at birth ranged from 67.77 years in Lithuania to 79.79 years in Italy (a difference of 12.02 years) and female life expectancy ranged from 77.37 years in Bulgaria to 85.32 years in Spain (a difference of 7.96 years), healthy life expectancies at birth ranged from 52.33 years in Slovakia to 71.68 years in Sweden (males – a difference of 19.35 years) and from 52.04 years in Slovakia to 71.64 years in Malta (females – a difference of 19.60 years). Thus the cross-national disparity in healthy life expectancy was considerably greater for both sexes than that in overall life expectancy. The analysis further split expectation of life with a disability into two subcategories – expectation with 'moderate activity limitation' and expectation with 'severe activity limitation'. Life expectancy at birth with moderate activity limitation ranged from 15.31 years for Slovenia to 4.71 years for Sweden (males – a threefold disparity) and from 18.80 years for Finland to 7.17 years for Sweden (females – well over a twofold disparity). Similarly life expectancy at birth with severe activity limitation ranged from 7.98 years for Slovenia to 1.88 years for Bulgaria (males) and from 10.18 years for Slovenia to 2.45 years for Bulgaria (females). The figures for Bulgaria are suspiciously low given its comparatively low overall life expectancies, unless of course people with severe activity limitation simply don't survive long in that country.

Disability-Adjusted Life Years (DALYs) and the Burden of Disease

Developed by Christopher Murray and Alan Lopez (Murray 1994; Murray and Lopez 1994, 1996; Lopez 2005), although with roots in earlier literature, the ***disability-adjusted life year***, or ***DALY***, has been adopted by WHO as a key public health metric combining morbidity and mortality in a single index. It has been an evolving metric, its elaborate methodology subject to constant review and fine tuning in response to the intense debate generated since its initial development in the early 1990s (see Murray (1996) and Murray et al. (2012a), including their substantial appendix). Thus not all studies based on DALYs define them in exactly the same way. The measure is designed to give effect to the concept of the ***burden of disease*** (BOD – sometimes extended to the burden of disease ***and injury***) in a population, either overall or from particular health conditions or health risk factors, or groups of conditions or risk factors. There have thus far been three Global Burden of Disease (GBD) studies, GBD 1990, GBD 2000 and GBD 2010. A summary of methodological developments across these studies can be found in the appendix to Murray et al. (2012a), but note that GBD 2010 made a point of generating comparable results for 1990 as well as 2010 (Lim et al. 2012; Murray et al. 2012b; Salomon et al. 2012b; Vos et al. 2012).

Burden of disease can be thought of as a measure of the gap between current health status and an ideal health status under which individuals universally live to old age free from disease and disability. It is promoted as an aid in setting health service and health research priorities and in identifying disadvantaged groups and targeting health interventions appropriately, and as providing a measure of output for intervention, program and health sector evaluation that facilitates ***legitimate comparison*** (before/after, through time and across populations). The quest that yielded DALYs as a measure had three main objectives. The first was to facilitate the proper incorporation of non-fatal health outcomes into health policy debate. Some of these outcomes, such as depression and musculoskeletal conditions, ranked far lower as health priorities than they should have when those priorities were set on the basis of mortality data alone. The second objective was to separate epidemiological assessment from advocacy by providing an objective measure of the mortality/disability arising from a health condition or risk factor. This was desirable to place in perspective the arguments of lobby groups with vested interests seeking funding to address particular health conditions or risk factors, and thereby avoid their advocacy skills dominating and distorting debate. The third objective was to produce a measure that could be used for both quantifying burden and analysing the cost-effectiveness of intervention options.

Broadly speaking, the number of DALYs for a disease, health condition or health risk factor is ***the sum of (i) years of life lost due to premature mortality caused by that disease, condition or risk factor (YLLs) and (ii) years of healthy life lost due to disability attributable to the disease, condition or risk factor (YLDs)***. The DALY thus extends the concept of potential years of life lost due to premature death, which it refines to also include years of 'healthy' life lost through being in poor health or disabled. One DALY is one lost year of 'healthy' life. In case you are wondering what exactly is meant by 'health risk factors', the ten globally most significant among 67 that were identified in GBD 2010 were, in rank order by number of lost DALYs, high blood pressure, tobacco smoking (including passive inhalation of second-hand smoke), household air pollution from solid fuels, diet low in fruits, alcohol use, high body-mass index, high fasting plasma glucose (indicative of diabetes), childhood underweight, ambient particulate matter pollution, and physical inactivity and low physical activity. This rank ordering varied (i) from those for males and females separately and (ii) from that for 1990 (Lim et al. 2012: Figures 1 and 2).

The methodology for calculating DALYs seeks to embrace four general concepts. First, ***as far as possible, any health outcome that represents a loss of welfare should be included as an indicator of health status***. The GBD 'cause list' identifies discrete conditions in a four-level hierarchy that begins at level 1 with three broad categories. These then subdivide into level 2, 3 and 4 subcategories. The base-level categories are (i) communicable, maternal, neonatal and nutritional disorders, (ii) non-communicable diseases and (iii) injuries. Level 2–4 subcategories have multiplied over time from 131 in GBD 1990 to 159 in GBD 2000 and 288 in GBD 2010. Beyond these subcategories 'sequelae' are also identified (i.e., direct consequences of disease or injury that are not otherwise captured in the cause list) –

483 of them in GBD 1990, 474 in GBD 2000 and 1,160 in GBD 2010 (Murray et al. 2012a: Appendix). Second, ***the characteristics of an individual affected by a health outcome that should be considered in calculating the associated burden of disease should be restricted to age and sex alone***. Third, ***like health outcomes should be treated alike*** – i.e., the premature death of a person of a given age and sex should contribute equally to the global burden of disease regardless of where, geographically, that person lives (Dhaka slum or salubrious Sydney suburb). This concept, advanced for GBD 1990, was modified for GBD 2010 by adjusting for 'comorbidity' – i.e., recognizing that some individuals' health loss has multiple contributing conditions, and that their health loss should be apportioned among those conditions rather than each condition being assessed separately, leading to exaggerated estimates of health loss at the individual level. Finally, ***the unit of measurement for the burden of disease should be time***. The second and third of these concepts in particular rule out discrimination by socio-economic status, whether in favour of the better off, on the ground that they contribute more to societal wellbeing, or in favour of the disadvantaged, on the ground that their very disadvantage justifies such favouritism. Murray (1996: 7) observes that they 'give DALYs a strongly egalitarian flavor.' Another debate concerning GBD methodology has been over ***whether incidence or prevalence measures*** (see Chap. 1) ***should be used in assessing disability from particular causes***. The initial position of Murray (1996) was that death rates were incidence rates and measures of prevalence made no sense in a mortality, or YLL, context. Therefore consistency recommended also using an incidence approach in performing YLD calculations. However, by GBD 2010 the debate had swung strongly in favour of taking a prevalence approach to measuring YLDs (Murray et al. 2012a: Appendix).

Five important 'social choices' are built into the construction of DALYs.

1. ***Estimating the duration of life lost due to a death at each age***. Since Dempsey (1947) first raised the issue of time-based measures of premature mortality, an array of possibilities have been suggested. Two of these were the alternative measures of YPLL discussed above – the one where life lost is determined with reference to a fixed upper cut-off age for deaths at all ages and the one where it is set for deaths at each age as the midpoint of the age group plus average remaining expectation of life at that age. Murray (1996) refers to these, respectively, as 'potential years of life lost' (PYLL) and 'period expected years of life lost' (PEYLL). A third option he considers is 'cohort expected years of life lost' (CEYLL), which differs from PEYLL only in that the life expectancies employed are from a relevant ***cohort*** life table rather than a standard, or period (cross-sectional) life table. The argument for this approach is that twentieth century experience is that cohort mortality beyond a given age rarely matches what a standard life table constructed at the time a cohort reaches that age predicts, because mortality in most populations has been declining (and hence survival prospects have been improving) over time. However, Murray rejects both PEYLL and CEYLL as violating the concept that like health outcomes should be treated alike, since life expectancies employed in respect of deaths

at a given age for a given sex will vary from population to population. Instead he opts for 'standard expected years of life lost', or SEYLL, in estimating the YLL component of DALYs. This approach uses ***an ideal standard life table*** to represent the level of survivorship corresponding to optimal health and to provide, ***for all populations of a given sex***, the remaining life expectancy lost through a premature death at each age. The approach ***allows deaths at the same age in all populations to contribute equally to the burden of disease in that population***, so that like outcomes are treated as like. For GBD 1990 the female and male standard schedules of remaining life expectancies were derived from Coale and Demeny 'West' model life tables for females with life expectancies at birth of 82.5 years and 80.0 years respectively. These were chosen on the basis that (i) the highest achieved life expectancy at birth at the time, for Japanese females, was over 82 years, and (ii) there was solid evidence that a comparable male life expectancy should be a little lower. The use of a female model life table to set the male standard schedule reflected the fact that no male model table with a life expectancy as high as 80 existed in the Coale and Demeny system at that time (although it was in the process of being added (Coale and Guo 1989)). By the time of GBD 2010, however, a single standard, or reference, life table with a life expectancy at birth of 86.0 years had been developed and was used for YLL calculations for both sexes (Murray et al. 2012a).

2. ***Comparing time lost due to premature death and time lived with a non-fatal health outcome***. The task of quantifying time lived with non-fatal health outcomes in a manner comparable to time lost due to premature mortality and standardized across populations and over time requires considerable simplification. While death is death and its measurement is relatively straightforward, non-fatal health outcomes are each unique in many ways and 'a certain level of reductionism must be accepted' (Murray 1996: 23). One genre of approaches to conceptualizing non-fatal health outcomes is a huge array of 'health-related quality of life' (HRQL) approaches. This genre was rejected as a basis for measuring the YLD component of DALYs on several grounds, but not least its substantial reliance on ***self-reports*** of health status. These have commonly yielded counter-intuitive patterns that suggest the measures in question do not meet the criterion of treating like outcomes alike in all populations. The alternative conceptualization that informs measuring YLDs is embedded in the ***International Classification of Impairments, Disabilities and Handicaps*** (ICIDH), a WHO classification that envisages a linear progression from 'disease' to 'pathology' to 'manifestation' to 'impairment' to 'disability' and then to 'handicap'. The key distinctions are between impairment (defined at the level of the organ system), disability (defined in terms of impact on the performance of the individual) and handicap (defined in the context of overall consequences, which depend on a person's social environment). Measurement of the YLD component of DALYs is ***focused on the measurement of disability***, not handicap (whose social environment element is incompatible with treating like outcomes alike), and also focused wherever possible on ***observations*** about non-fatal health

outcomes rather than self-reports (although the latter are the only option for outcomes categorized under 'pain and suffering').

The process for calculating YLDs is complex and has undergone significant revision with each new GBD study in light of debate generated by previous studies. Central to it is the determination of ***disability weights*** for each non-fatal health state that effectively rank those health states from least to most disabling on a scale ranging from 0 (no disability – perfect health) to 1 (complete disability – death). Murray et al. (2012a: Appendix, p. 11) describe disability weights as 'the key mechanism in the GBD approach through which disease and injury sequelae are made comparable with each other and with time lost due to premature mortality.' For GBD 1990, after an initial methodology had attracted criticism (Murray 1996: 34), weights were established using a deliberative process that first asked a group of health professionals to participate in an exercise to achieve a consensus ranking of 22 'indicator conditions' chosen to reflect different dimensions of non-fatal health outcomes by their disabling severity. 'Dimensions' traversed conditions with largely physical manifestations such as blindness and below the knee amputation, neuro-psychiatric conditions, conditions representing varying degrees of pain, conditions that affect sexual/reproductive function, and one condition, vitiligo on the face, that has exclusively social or group interaction consequences (vitiligo causes depigmentation of parts of the skin, thereby affecting physical attractiveness). The ranking obtained was then used to establish seven 'disability classes', each associated with a discrete range of severity, or disability, weights (Murray 1996: Table 1.4) ranging from class 1, weight range 0.00–0.02 (indicator conditions vitiligo on the face, weight-for-height lower than two standard deviations below the mean) to class 7, weight range 0.70–1.00 (indicator conditions active psychosis, dementia, severe migraine, quadriplegia). Disability weights for hundreds of other conditions were then agreed on by expert panels with reference to these disability classes and the weight ranges associated with them (what class did a condition belong in, then where within the associated range of disability weights did it fit?).

For GBD 2010 disability weights were comprehensively revised in the wake of considerable debate about the GBD 1990 weights. The new method (Salomon et al. 2012a) clarified its objective as being to measure ***health loss rather than welfare loss*** and moved to using ***the general public*** (rather than health professionals or the third option, individuals actually in a particular health state) to make the comparative assessments required to establish disability weights. The 1,160 disease and injury sequelae referred to above were mapped into 220 distinct 'health states' capturing the most salient differences in symptoms and functioning, and household surveys in five deliberately contrasting countries – Bangladesh, Indonesia, Peru, Tanzania and the USA – were used, along with a web-based survey posted online for almost 9 months, to elicit responses on ***paired comparison questions***. These presented respondents with descriptions of two hypothetical people, each with a particular health state, and asked them which person they deemed healthier. In the household surveys each respondent was asked to respond in respect of 15 paired comparisons randomly selected from a pool of 108 health states. These were chosen from the broader pool of 220 on the basis that (i) their salient features could be easily communicated to lay respondents in 35 or fewer words, and (ii) they could be presented realistically as chronic states that would persist throughout a lifetime. Respondents to the web-based survey were randomly assigned to one of four survey versions, one of which mimicked the household survey while the others focused on different subsets of the 220 health states. One of these was a subset of 30 health states in respect of which respondents were asked to compare the health benefits of different life-saving or disease-prevention programs as an aid to anchoring disability weights to the necessary zero to one scale. There were 13,902 participants in the five household surveys and 16,328 in the web-based survey. The complex analysis of these data to generate disability weights with 95 % 'uncertainty' intervals about them (another new innovation compared to GBD 1990) is described in Salomon et al. (2012a), where the weights for the 220 health states are also presented. Health states with the highest disability weights were acute schizophrenia (0.756), severe multiple sclerosis (0.707), untreated spinal cord lesion at the neck (0.673), severe epilepsy (0.657), severe major depressive disorder (0.655), heroin and other opioid dependence (0.641), severe traumatic brain injury with long-term consequences (0.625) and severe musculoskeletal problems (0.606). On the other hand 26 % of all health states had disability weights below 0.05, and 11 % had weights below 0.01. These were obviously minimally disabling states.

3. ***Discounting future health***. Discounting is an economic concept that reflects the fact that most people prefer benefits now rather than in the future. There has been considerable debate among health economists over whether future health should be discounted in the calculation of YLLs and YLDs, and hence of DALYs. For summaries of this debate see Murray (1996) and Murray et al. (2012a: Appendix). In Murray et al.'s view the more persuasive arguments in favour of discounting future healthy life years focus upon decision-making paradoxes that without discounting invite unpalatable choices. In particular, the 'research and eradication' paradoxes that, in the absence of discounting, suggest that spending money on research and eradication could yield extremely large streams of healthy life years in the future, therefore justifying devoting nearly all resources ***today*** to those areas, at the expense of properly treating those currently suffering from a health condition. Thus, the arguments for discounting are not arguments that health in the future should intrinsically be valued less than health in the present. They are arguments that one should discount to avoid unreasonable sacrifice by those currently living with a condition.

   For GBD 1990 a 3 % discount rate was adopted, but vigorous debate ensued subsequently that by the time of GBD 2010 had generated a consensus view that a year of healthy life should be deemed equally important to population health, regardless of the year in which it was lived. Hence GBD 2010 used a zero discount rate. The important thing to bear in mind is that whether or not one discounts makes a non-trivial difference to YLL, YLD and DALY results. A 3 % compounding discount rate makes a big difference by the time one gets to older ages for someone who dies in infancy, and significantly alters the relative importance attached to deaths in childhood compared to those in adulthood in BOD studies.

4. ***Age-weighting***. This became an issue for GBD studies on the basis that empirical evidence from a range of sources showed that people tend to assign greater importance to saving the lives of adolescents and young adults than to saving those of very young children and older adults (Murray 1996). In other words, there is a tendency to value life-years lived at different ages differently. This may have a variety of rationales for different individuals, but it is likely to reflect such things as the greater value to society of years lived through the working ages, when a return is obtained on the investment in a person's education and people are productive economically and actively or potentially involved in producing and raising the next generation, than of those lived through childhood and old age in varying degrees of dependence. As a result of these considerations GBD 1990 built into the computation of DALYs age weights that assigned greater value to a year of young or middle-aged adult life than to a year lived as a young child or one of the elderly. These weights were determined via a continuous mathematical function that increased from a value of 0 at birth to a peak of around 1.5 at ages in the mid-20s, then tailed off thereafter to around 0.5 at age 80, and less at still older, more dependent, ages.

   Yet again, however, methodological debate led to simplification of the calculation of YLLs, YLDs and DALYs for GBD 2010. Viewing these as strict summary measures of population health it was eventually agreed that arguments for weighting years of healthy life lived at different ages differently were not as compelling as previously thought. In consequence it was concluded that 'we should treat a year of healthy life as equal irrespective of the age at which it is lived' (Murray et al. 2012a: 2064). In other words, no age weighting was employed in GBD 2010.

5. ***Considerations of equity***. In developing the methodology for GBD 1990 there was consideration given to a couple of issues of equity. Empirical evidence has suggested that people generally have a preference for a smaller number of people to receive a substantial health benefit than for a larger number of people to receive a smaller one. The possibility was explored of incorporating this sort of distributional preference into DALY methodology, but it was deemed not to be justified. Another argument has been that DALYs should somehow be weighted socio-economically; that life-years for the disadvantaged should be more heavily weighted than life-years for the better off. This proposition, too, was rejected after consideration, it being concluded that the needs of disadvantaged groups could be met simply by studying and reporting on differential burden patterns by socio-economic status. Neither of these issues appears to have been seriously revisited for GBD 2010.

References

Albert, F. S., Bragg, J. M., & Brooks, J. C. (2008). *Health expectancy*. Paper presented at the Living to 100 and Beyond Symposium, Orlando, Florida.

Brass, W. (1971). On the scale of mortality. In W. Brass (Ed.), *Biological aspects of demography* (pp. 69–110). London: Taylor and Francis.

Bruggink, J.-W. (2011). *Towards a better health expectancy*. The Hague: Statistics Netherlands.

Chiang, C. L. (1968). *Introduction to stochastic processes in biostatistics*. New York: Wiley.

Chiang, C. L. (1984). *The life table and its applications*. Malabar: Robert E. Krieger Publishing Company.

Coale, A. J., & Demeny, P. G. (1966). *Regional model life tables and stable populations*. Princeton: Princeton University Press.

Coale, A., & Guo, G. (1989). Revised regional model life tables at very low levels of mortality. *Population Index, 55*(4), 613–643.

Coale, A.J., Demeny, P.G., & Vaughan, B. (1983). *Regional model life tables and stable populations* (2nd ed.). New York/London: Academic.

Dempsey, M. (1947). Decline in tuberculosis. The death rate fails to tell the entire story. *American Review of Tuberculosis, 56*, 157–164.

Fergany, N. (1971). On the human survivorship function and life table construction. *Demography, 8*(3), 331–334.

Flieger, W., & Cabigon, J. (1994). *Life table estimates for the Philippines, its regions and provinces, by sex: 1970, 1980 & 1990*. Manila: Department of Health and USAID.

Gardner, J. W., & Sanborn, J. S. (1990). Years of potential life lost (YPLL) – What does it measure? *Epidemiology, 1*(4), 322–329.

Greville, T. N. E. (1943). Short methods of constructing abridged life tables. *Record of the American Institute of Actuaries, 32*(65), 29–42.

Guillot, M., & Yu, Y. (2009). Estimating health expectancies from two cross-sectional surveys: The intercensal method. *Demographic Research, 21*, 503–534.

Keyfitz, N. (1970). Finding probabilities from observed rates or how to make a life table. *The American Statistician, 24*(1), 28–33.

Keyfitz, N., & Frauenthal, J. (1975). An improved life table method. *Biometrics, 31*, 889–899.

Lim, S., Vos, T., Flaxman, A. D., & 206 others. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*, 2224–2260.

Lopez, A. D. (2005). The evolution of the global burden of disease framework for disease, injury and risk factor quantification: Developing the evidence base for national, regional and global public health action. *Globalization and Health*. doi:10.1186/1744-8603-1-5.

Mathers, C., McCallum, J., & Robine, J.-M. (1994). *Advances in health expectancies*. Canberra: Australian Institute of Health and Welfare.

Murray, C. J. L. (1994). Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bulletin of the World Health Organization, 72*(3), 429–445.

Murray, C. J. L. (1996). Rethinking DALYs. In C. J. L. Murray & A. D. Lopez (Eds.), *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020* (pp. 1–98). Boston: Harvard University Press for the Harvard School of Public Health, WHO and the World Bank.

Murray, C. J. L., & Lopez, A. D. (1994). Quantifying disability: Data, methods and results. *Bulletin of the World Health Organization, 72*(3), 481–494.

Murray, C. J. L., & Lopez, A. D. (1996). *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Boston: Harvard University Press for the Harvard School of Public Health, WHO and the World Bank.

Murray, C. J. L., Salomon, J. A., Mathers, C. D., & Lopez, A. D. (2002). *Summary measures of population health: Concepts, ethics, measurement and applications*. Geneva: WHO Press.

Murray, C. J. L., Ferguson, B. D., Lopez, A. D., Guillot, M., Ahmad, O., & Salomon, J. A. (2003). Modified logit life table system: Principles, empirical validation and application. *Population Studies, 57*(2), 165–182.

Murray, C. J. L., Ezzati, M., Flaxman, A. D., Lim, S., Lozano, R., Michaud, C., Naghavi, M., Salomon, J. A., Shibuya, K., Vos, T., Wikler, D., & Lopez, A. D. (2012a). GBD 2010: Design, definitions and metrics. *The Lancet, 380*, 2063–2066. Includes a substantial appendix by the same authors separately titled ‘Comprehensive systematic analysis of global epidemiology: Definitions, methods, simplification of DALYs, and comparative results from the Global Burden of Disease Study 2010’.

Murray, C. J. L., Vos, T., Lozano, R., & 358 others. (2012b). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*, 2197–2223.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modelling population processes*. Oxford: Blackwell.

Reed, L. J., & Merrell, M. (1939). A short method for constructing an abridged life table. *American Journal of Hygiene, 30*(2), 33–62.

Robine, J.-M., Romieu, I., & Cambois, E. (1999). Health expectancy indicators. *Bulletin of the World Health Organization, 77*(2), 181–185.

Robine, J.-M., Jagger, C., Mathers, C. D., Crimmins, E. M., & Suzman, R. M. (2003). *Determining health expectancies*. Chichester/Hoboken: Wiley.

Salomon, J. A., Vos, T., Hogan, D. R., & 125 others. (2012a). Common values in assessing health outcomes from disease and injury: Disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet, 380*, 2129–2143.

Salomon, J. A., Wang, H., Freeman, M. K., Vos, T., Flaxman, A. D., Lopez, A. D., & Murray, C. J. L. (2012b). Healthy life expectancy for 187 countries, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*, 2144–2162.

Sanders, B. S. (1964). Measuring community health levels. *American Journal of Public Health, 54*(7), 1063–1070.

Shryock, H. S., Siegel, J. S., and Associates. (1973). *The methods and materials of demography*. 2nd Printing (revised). Washington, DC: U.S. Bureau of the Census.

Siegel, J. S., & Swanson, D. A. (2004). *The methods and materials of demography* (2nd ed.). San Diego/London: Elsevier Academic Press.

Smith, D. P. (1992). *Formal demography*. New York/London: Plenum Press.

Sullivan, D. F. (1971). A single index of mortality and morbidity. *HSMHA Health Reports, 86*(4), 347–354.

United Nations. (1955). *Age and sex patterns of mortality: Model life tables for under-developed countries*. New York: United Nations.

United Nations. (1982). *Model life tables for developing countries*. New York: United Nations.

Vos, T., Flaxman, A. D., Naghavi, M., & 354 others. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*, 2163–2196.

World Health Organization. (1977). *Manual of mortality analysis: A manual on methods of analysis of national mortality statistics for public health purposes*. Geneva: WHO.

Zhao, Z. (2007). Interpretation and use of the United Nations 1982 model life tables: With particular reference to developing countries. *Population-E, 62*(1), 89–116.

Chapter 5
Marriage, Marital Status and Relationships

Why Study Marriage?

Unlike births and deaths, the two types of events of primary interest to demographers, marriage is not a biological event but a ***social event***. Historically it has been of interest to demographers principally because of its close connection with the process of fertility. In most societies marriage has traditionally signalled the commencement of regular sexual activity and hence of serious exposure to the risk of childbearing. Equally, provided the biological procceses associated with getting older have not already done so, ***marriage dissolution***, whether through divorce or widowhood, tends (sometimes only temporarily) to terminate exposure to the risk of childbearing.

This is not, of course, to suggest that unmarried people are not sexually active and do not have children. Indeed, in many of the populations of Europe, North America and Australasia non-marital sexual activity and childbearing have increased drammatically over recent decades as the second demographic transition (Van de Kaa 1987) has seen consensual partnering proliferate, greatly weakening the link between formal marriage and fertility and raising questions as to the definition of 'marriage'. The link between marriage and childbearing is also less than complete in many African societies. However, it is still true to say that the great majority of fertility in most populations is ***marital*** fertility, as distinct from ***non-marital***, or ***ex-nuptial***, or, as it was often pejoratively labelled historically, ***illegitimate*** fertility.

It is often the case when studying changes in fertility levels and patterns over time that a major part of the explanation for those changes is to be found in concurrent changes in marriage patterns. People have been getting married earlier or later in life (i.e., at younger or older ages), and/or proportionately more of them or fewer of them have been marrying at all. In Australia, for example, there was considerable alarm over declining fertility in the late nineteenth and early twentieth centuries. Much of that decline was due to later and less universal marriage, and to consequent

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_5

diminished exposure to risk. Likewise the well-known post-war baby boom in Australia has been attributed primarily not to increased fertility within marriage, but to women marrying younger and to proportionately fewer of them failing to marry at all. In developing countries with high fertility levels, some of the most effective strategies for reducing fertility have been strategies which sought to delay entry into marriage.

Defining Marriage

A ‘marriage’ is defined by the *Multilingual Demographic Dictionary* (IUSSP 1958) as a ‘union between persons of opposite sexes which involves rights and obligations fixed by law or custom’. Another view is that marriages are legally recognized unions between a man and a woman, recognition being accorded under one or more of ***religious*** law, ***civil*** law or ***customary/tribal*** law. In most societies a couple marries by passing through a ceremony or ritual, prior to which they are not married but after which they are. There are, however, also societies in which becoming married is more a process than an event, so that it is not necessarily easy to date precisely when the married state was entered, or indeed to say whether or not it has been reached at a given point in time. Indigenous Australians are a case in point, and the marriage practices of some African societies also have been claimed to fit this type of model.

In addition to marriages demographers also are interested, and increasingly so these days in developed countries in particular, in ***consensual unions*** (also referred to by a variety of other labels, including ***cohabiting unions***, ***living together relationships***, ***de facto marriages*** and ***de facto relationships***). These are unions of cohabiting heterosexual partners in which formalities signifying marriage in the society in question have not been completed. They tend to vary greatly in the degree of mutual commitment of the cohabiting parties to one another, and therefore in the stability of the unions themselves. At one extreme they can be marriages in all but law. Sometimes one or both parties is unable to marry because of inability to legally dissolve a previous marriage, while on other occasions formal marriage is viewed as adding nothing to the relationship, as complicating its dissolution should the need arise, as attaching to it attributes (e.g., sex role stereotypes; a subservience of the female to the male partner) that are viewed negatively, as something over which there is no particular urgency, etc. At the other extreme consensual unions are characterized by levels of mutual commitment that are decidedly less than marriage-like. At the lowest level they may be little more than convenient settings for regular sexual gratification, exploitation of a woman by a man for the domestic services she can provide, and/or minimising living costs. At a somewhat higher level they may be essentially ‘going steady’ relationships in which couples enjoy one another’s company and intimacy (and the optimal conditions for this afforded by cohabitation), but are either not seriously interested in long-term commitment at that time in their lives or intent on thoroughly testing the relationship before contemplating making such a commitment. Consensual unions of this type have a proneness to dissolution that derives from inherent risks that either they will ultimately be deemed to have failed the test by one or both parties, or the parties will not progress to a desire for longer term commitment at the same pace.

Consensual unions are not infrequently ***precursors*** to marriage, but as just intimated may also be ***alternatives*** to marriage or ***devoid of marital character***. Individual unions may merit classification into more than one of these categories over time, or in the perceptions of the two parties to them at a point in time, and in circumstances of ready access to fertility control this type of categorization is likely to have implications for levels of childbearing. In consensual unions that are perceived to be informal marriages the propensity to bear children may differ little from that within formal marriages. In contemporary Sweden, for example, formal marriage has virtually no role as a legitimator of parenthood, and a high proportion of essentially marital childbearing takes place within consensual unions. One does, however, need to remember that in some countries significant proportions of informal marriages are associated with repartnering after marriage breakdown and with decisions to remain voluntarily childless, both of which tend to depress fertility. Where consensual partnering is a precursor to marriage the propensity to have children may also be quite high, and indeed pregnancy may tend to be associated with the transition to marriage, either as a trigger or as part of a planned dual change of status. On the other hand, those in relatively uncommitted consensual unions probably tend to view that phase of their lives as ideally childless, although clearly women have more incentive to ensure that it turns out that way than do men.

Recent decades have also seen recognition accorded to ***LAT relationships***. LAT stands for ***living apart together***, and describes couples who are in intimate relationships that are not co-residential. Reasons for this can be varied. They may be pragmatic, as when differing work locations force residential separation, or a couple (perhaps both living with parents) can't afford to co-reside, or a single mother would jeopardize her entitlement to a welfare benefit were she to co-reside with a boyfriend who was earning an income. They may be ideologically based, as when couples or individual members thereof value the autonomy separate residence provides in conjunction with the intimacy they enjoy. For others they may represent a stage in the process of developing a relationship – a precursor to possible cohabitation and subsequently even marriage. Then again, a party to a broken relationship who has custody of its children may wish to keep a subsequent romantic interest at arms length from those children, or a previous unpleasant experience of cohabitation may promote wariness, at least in the meantime, of establishing co-residence with a new intimate partner. Motivations for LAT relationships can be complex, and LAT couples can be variously gladly, regretfully or ambivalently living apart. The fact that these relationships are intimate, however, obviously gives them potential to contribute to fertility, although in most societies that potential is probably less likely to be realized than is the potential in co-residential intimate relationships.

A particular variant of LAT relationships which is of less recent origin and in which childbearing is common is what are known as ***visiting unions*** (sometimes also termed ***extra-residential unions***). They are widespread in Caribbean societies among majority groups of African slave (as opposed to Indian or Chinese) descent. Visiting unions exist alongside common-law (consensual) unions and formal marriages, and relationships occasionally proceed through all three forms as a man's capacity to economically support his family increases. They are, however, especially common at lower socio-economic levels, are associated with ***matrifocal family structures*** (female breadwinners who often join forces with female kin, especially mothers and sisters, to share income generation, childcare and domestic duties), and have helped generate an extensive literature seeking to understand and explain Caribbean mating and family formation patterns (Barrow 1996, 1998). Characteristics attributed to visiting unions include: (i) youthful commencement; (ii) variable, but commonly short-lived durations; (iii) a depressant effect on fertility (earlier, but more widely spaced and rapidly ended childbearing compared to marital unions); (iv) lack of social obligations, which facilitates easy termination by either party; (v) conducive to frequent partner changing and multiple partnering (often, from women's perspective, from necessity rather than desire); and (vi) the emphasis they give to mother-child, and especially mother-daughter, relationships. Among rationales offered for the prominence of visiting unions in the Caribbean, where males have considerable sexual freedom and all except young teenage girls can also participate in premarital mating without community censure, have been: (i) the ease with which they can be managed because a lack of formal kinship ties and expectations maximizes personal choice; (ii) their being a rational adaptive response to chronic poverty, unemployment and underemployment that renders fulfilling conventional marital economic roles impossible for lower class males; (iii) the flexibility they offer both sexes to respond quickly to both socio-economic constraints and economic opportunities (often through migration) as and when they arise; (iv) their offering women personal autonomy, shelter from conjugal violence and scope for 'child shifting' (i.e., transferring dependent children to households not including a natural parent, usually in the interest of economic survival); and (v) the attraction of childbearing within one's mother's home before tackling the potential difficulties of a co-residential union. In short, Caribbean visiting unions are in varied ways ***a coping mechanism*** for those of African descent at lower socio-economic levels. Rubenstein (1980: 333) writes of them, as observed in a village on the island of St Vincent, as follows:

> The female is expected to be sexually faithful to her partner, who, in turn, is obliged to reciprocate with regular or periodic gifts. Beyond this, little is specified about the way the union should be acted out .... Various types of behaviour are therefore permitted. Some women regularly wash and cook for their non-resident boyfriends, while others never do so. Some unions are very intimate, while others are very formal and confined to sexual release. Some unions involve constant visits between households, while others are based on brief encounters in deserted spots or in the home of a third party. Some unions require considerable economic assistance from the male, while others are characterized by small or infrequent gifts. There is a measure of male sexual exclusiveness in certain unions; others represent an expression of the male desire to have concurrent affairs with several women. Some show promise of being a prelude to marriage; others are simply fleeting affairs soon to be terminated.

The phrase ***conjugal unions*** is sometimes used to refer to the combination of formal marriages and consensual unions; i.e., to all stable co-residential sexual unions. The term ***marital unions*** can have the same meaning, but may also be used to refer only to formal marriages, so when encountered its precise meaning should be checked. Demographers ideally prefer to conduct analyses of all conjugal unions, but data sources available to them often preclude this. It is important to realize also that different data sources may define 'marriage' differently.

In many countries the marriages of couples passing through recognized religious or civil marriage ceremonies are required to be registered with a government agency, and the regularly released ***registration data*** yielded by this administrative procedure are widely used in the analysis of marriage patterns and trends. These data concern marriage ***events***, and adopt a particular, legalistic definition of marriage. Their availability also means that marriages are more easily studied than are consensual unions, data on which tend to have to be obtained from special surveys. ***Census data*** on marriage do not pertain to marriage events. They typically focus on the status of individuals as married, or not married, at the time of the census count, and rely on respondents' perceptions of which category they fit into. There is scope for persons who are legally married to deny that they are, but more importantly there is scope for persons to claim to be married who have not taken the steps that would result in their being registered as such. This can occur because ritual and the registration process are seen as irrelevant to the nature and quality of the union, or because the notion of registered marriage is culturally alien. If the latter situation prevails for the bulk of a population there are unlikely to be any registration data available, and census data should reflect prevailing cultural norms and practices as to what constitutes marriage. Sometimes, though, registered marriage is the standard model within a majority culture, but is eschewed by a minority subculture. New Zealand's Maori and Australia's Indigenous minorities are two examples. Both, in the eyes of dominant European cultures, are particularly prone to form consensual unions, yet many of these unions are marriages according to Maori or Indigenous custom, and are likely to be returned as such at a census. The result is that the census records significantly more married persons than registration data would lead one to expect, and a blurring of the distinction between marriages and consensual unions occurs. Should a union which is a marriage according to the cultural customs of the parties involved be labelled 'consensual' because those customs happen not to require registration? Mind you, with the proliferation of consensual unions in recent years some censuses have made attempts to enumerate such unions. The 2011 Australian census, for example, gathered such data through questions on the relationships between household members which distinguish between 'husband/wife' and 'de facto partner' relationships. In New Zealand's 2006 census individuals were asked to indicate who lived in the same household as they did, with two options being 'my legal husband or wife' and 'my opposite-sex partner or de facto, boyfriend or girlfriend'.

A third source of data on marriage is ***sample surveys***. If utilising a similar question focused on current status, these may impose the same definition of marriage as a census. However, demographic sample surveys often are at pains to

date events like marriages, and this tends to facilitate the separation of registered marriages from informal unions that, given the chance, respondents would claim to be 'marriages without the piece of paper'. Whether it also facilitates the separation of registered marriages from unions that are marriages by some other cultural prescription depends on whether that prescription lends itself to specification of a precise date of marriage. If it does, no distinction is likely to be possible unless the survey includes other questions for the purpose. However, if marriage is a state into which couples ***drift*** as consensual unions gradually are accorded community recognition as marriages, fixing a date of marriage is likely to be problematic and unions are likely to be deemed consensual. Such a procedure is open to the criticism that it demeans a particular cultural perspective on marriage.

Some Other Features of Marriage Systems

In many cultures marriage is not only a major social and demographic event, but a significant ***economic*** one too. In addition to religious and/or customary ritual it may entail non-trivial economic transactions. ***Dowry*** systems require economic transfers from the family of the bride to the groom and possibly also his family; ***bridewealth*** or ***bride-price*** systems require the groom to make payment to the family of the bride.

Where law or custom prescribe ***monogamy***, no person may have more than one husband or wife at any given time. The term ***serial monogamy*** is applied to situations where individuals experience a succession of monogamous relationships. Monogamous marriage systems make life much easier for the demographic analyst than the alternative polygamous systems. ***Polygamy*** is the social institution that permits a person to have multiple spouses at one time. Where men are permitted to have multiple wives the more specific social institution that is said to exist is ***polygyny***. The much rarer institution that allows women to have multiple husbands is ***polyandry***. Polygynous marriage systems, common in Africa and formerly widespread in Asia as well, have tended to be sustainable under conditions of high fertility, and hence very youthful age structures, in conjunction with marked age differences between husbands and their younger wives. Their viability is seriously threatened by marked fertility decline.

Demographers sometimes are interested in the extent to which particular groups of individuals tend to marry across social boundaries. ***Endogamy*** is said to exist when both parties to a marriage are from the same a tribe, clan or other similar social group; ***exogamy*** is the term used to describe marriages across such boundaries. Another similar distinction is that between ***homogamy*** and ***heterogamy***. Marriages are said to be homogamous when contracted between parties with similar social, physical or mental characteristics. Thus there might be interest in the extent to which marriages were religiously (social), racially (physical) or educationally (mental) homogamous. Marriages which entail intermarriage between members of different religious, racial, educational, etc. groups are said to be heterogamous.

Marital Status

'Marital status' is the term used to describe the categories into which the processes of marriage and marriage dissolution, collectively referred to as the processes of ***nuptiality***, transfer individuals. In all societies where formal marriage in some guise takes place there are at least three categories of marital status – ***never married***, ***married*** and ***widowed*** – the latter comprising persons whose marriages ended because their spouses died and who have not remarried. Where it is possible under civil, religious or customary law to formally terminate a marriage other than by one spouse dying there is also a category ***divorced***. It may not, however, because of legal requirements, be possible to enter this category immediately upon deciding to end a marriage. Thus a category ***separated***, or sometimes ***legally separated***, may exist to capture persons in transition from being married to being divorced. Members of this category remain legally married, and may reconcile with their estranged spouses and rejoin the 'married' group. It is sometimes, though, useful to be able to exclude them from the married population. In studying fertility, for example, unless already pregnant at the time of separation they are more likely to contribute to non-marital (or in the official Australian terminology, 'ex-nuptial') than to marital (or 'nuptial') childbearing, through having formed sexual relationships with men other than their husbands. Indeed, in a proportion of cases the formation of such relationships will have precipitated separation. In recent times census and survey takers have also, in recognition of widespread marital disruption and repartnering, sometimes sought to split the 'married' group into those in ***first marriages*** and ***remarriages***. Then again, for some purposes it is common for all groups except the never married to be aggregated as those who have ***ever married***.

It has already been noted that the 'marriage' data censuses typically gather are data on a population's marital status, and that they generally accept respondents' perceptions of the marital status categories to which they belong. Surveys may take the same approach, but have the capacity to probe more deeply and need not be as problematic. It is important to realize that the marital status distributions that result from asking 'What is your marital status?' and inviting respondents to tick a box corresponding to the category they belong to ***need not result in a classification based on a strictly legal definition of marriage***. Some cohabiting individuals may consider their relationships to be marriages according to another set of cultural prescriptions besides that enshrined in law, or simply on the basis of the levels of mutual commitment they feature. They may thus claim to be married even though, legally, they are not. The same can happen with the statuses 'separated', 'divorced' and 'widowed', with which persons might identify on the basis that a longstanding ***consensual*** union rather than a legal marriage had dissolved. In other words, census and survey marital status distributions may reflect ***social***, as well as ***legal***, definitions of marital status. Not even specifically asking for ***legal*** marital status guarantees that that is what respondents will give, and providing additional categories for the consensually partnered and those who have had consensual unions dissolve is no solution either. Those who might identify with such categories also have legal marital statuses with which they might identify, so to what extent do they opt for each alternative? The best that can usually be done is to ask ***separately*** about legal marital status and about living situation (where two response options might be 'Living with legal husband/wife' and 'Living with a partner unmarried'), but even this is not foolproof. If respondents wish to represent consensual unions as marriages, they will do so.

The issue just discussed can pose real problems for demographic analysis when ratio-type measures refined by marital status draw their numerators and denominators from different data sources. The classic example is the calculation of marital and non-marital fertility rates. The ***numerators*** of these rates normally come from birth ***registration*** systems, which classify as 'marital' only births to women legally married to the fathers of their children; i.e., the mother and father must be able to provide their date of marriage. The ***denominators***, however, typically come from ***census*** data, and to the extent that consensually partnered women claim to be 'married' they contribute to the denominator of the ***marital*** fertility rate, whereas births they have contribute to the numerator of the ***non***-marital fertility rate. One can argue over whether the births or the women are misclassified, but clearly both should be reflected in one measure; they should not be split between the two. The result is that the non-marital fertility rate is inflated (because its denominator excludes some of the women contributing, or potentially contributing, to its numerator) while the marital fertility rate is artificially depressed (because some of the women in its denominator cannot contribute births to its numerator). The picture can become really distorted if one is dealing with a group such as the Maori of New Zealand, to whom the system of registered marriage underpinning 'legal' marriage in the country is culturally alien. Many women who claim to be married at a census because they are married according to Maori custom may have their births registered as 'ex-nuptial' because Maori marriage does not yield a precise date of marriage. The result can be inflated non-marital fertility rates, and misleadingly low marital fertility rates.

The demographer's main interest in marital status derives from the different levels of opportunity for, and social acceptability of, childbearing associated with different categories of marital status. Changes in sex-age-specific distributions of the population by marital status may signal, or explain, changes in fertility. For example, age-specific marital fertility rates may remain unchanged over a period, but age-specific proportions married might fall; the result would be a decline in ***overall*** age-specific fertility rates, and therefore a decline in ***total*** fertility. A fundamental question that should always be asked when faced with a rise or fall in fertility is 'How much of the change is due to change in age-specific proportions married, and how much is due to change in fertility within marriage?'

General Measures of the Marriage Process

Predictably, the most basic measure of the intensity of the process of marriage is a crude rate, the ***crude marriage rate***. This was previously (Chap. 1, Eq. 1.8) defined as follows:

$$\text{CMR} = (\text{M}/\text{P})\ 1{,}000 \tag{5.1}$$

Where M = marriages during year y; P = mean (or mid-year) total population in year y.

Like other crude rates the CMR has major deficiencies, and is for ***cautious*** use when nothing more refined can be calculated. There is an argument that its numerator should be doubled to recognize the fact that there are two parties to each marriage, but this makes no difference to the relative levels of its value for different populations. The real problem is that it is not a true rate; its denominator includes individuals who cannot possibly contribute to its numerator – they are too young or, assuming the marriage system is monogamous, already married.

An initial refinement is to calculate a ***general marriage rate*** in which, as was the case with the general fertility rate, the denominator is a measure of the true population at risk, or the ***marriageable*** population. Such a measure excludes persons under the legal minimum age for marriage and those older persons who are already married (and not either widowed or divorced). It is usually suggested that the GMR should be calculated separately for males and females, but it is not a measure that is in common use. Its equation is:

$$\text{GMR} = (\text{M}/\text{P}_\text{m})\ 1{,}000 \tag{5.2}$$

Where M = marriages during year y; P_m = the mean (or mid-year) marriageable population of a given sex.

What ***are*** in common use are ***sex-age-specific marriage rates***. Using life table-type notation these are given by:

$$_\text{n}\text{SASMR}_\text{x} = ({}_\text{n}\text{M}_\text{x}/{}_\text{n}\text{P}_\text{x})\ 1{,}000 \tag{5.3}$$

Where x = the lower limit of the age group for which the marriage rate is being calculated; n = the width of that age group in years; ${}_\text{n}\text{M}_\text{x}$ = the number of marriages of males or females aged between x and x + n years (exact ages) during year y; ${}_\text{n}\text{P}_\text{x}$ = the mean (or mid-year) male or female population aged between x and x + n years (exact ages) in year y.

These rates, however, are in the same category as the CMR in having denominators which include persons who are already married and therefore not at risk of marrying during year y. To refine them to overcome this deficiency we can calculate ***sex-age-marital status-specific marriage rates***. These are given by:

$$_{n}\mathrm{SAMSMR}_{x} = ({}_{n}\mathrm{M(m)}_{x}/{}_{n}\mathrm{P(m)}_{x})\ 1{,}000 \tag{5.4}$$

Where x = the lower limit of the age group for which the marriage rate is being calculated; n = the width of that age group in years; $_{n}\mathrm{M(m)}_{x}$ = the number of marriages of males or females of marital status m aged between x and x + n years (exact ages) during year y; $_{n}\mathrm{P(m)}_{x}$ = the mean (or mid-year) male or female population of marital status m aged between x and x + n years (exact ages) in year y.

Where the marital status m is 'never married', Eq. 5.4 yields sex-age-specific ***first*** marriage rates. Where m is 'widowed' or 'divorced' it yields sex-age-specific ***re***marriage rates. The latter should be calculated separately for widowed and divorced persons, not for the two marital statuses combined, because levels of remarriage tend to be quite different depending on which mechanism (death or divorce) terminated the previous marriage. Should the data on marriages classified by marital status required to apply Eq. 5.4 not be available, one option to improve on SASMRs as defined in Eq. 5.3 is to restrict their denominators $_{n}\mathrm{P}_{x}$ to the mean (or mid-year) populations who were never married, widowed or divorced and aged between x and x + n years (exact ages). This yields 'true' rates by eliminating persons not at risk from the denominators. One lesson from having presented this option is that you should always, when encountering SASMRs, check what denominator was used in their calculation – was it a total population denominator or a 'not currently married' population denominator?

First Marriage

Reference to first marriage rates and remarriage rates introduces a fundamental refinement of any detailed analysis of marriage patterns and trends. You may recall that in Chap. 1 we distinguished between renewable and non-renewable events. In Chap. 3 we noted that renewable events (like marriage) posed problems for demographic analysis, especially in the construction of probabilistic types of measures. However, these problems could be overcome by subdividing a renewable event process into several ***non***-renewable processes. ***Any detailed study of marriage MUST look separately at first marriages*** (which are non-renewable events), and ideally should also break subsequent marriages down by ***order*** (second, third, etc. marriages) and whether remarriage followed widowhood or divorce.

Measures of the Intensity of the First Marriage Process

One of the more commonly encountered measures of the intensity of the process of first marriage these days, especially in the work of European demographers, is the

total first marriage rate, or TFMR. As its name suggests, it is directly analagous to the more familiar total fertility rate, and is constructed in a similar way. It is a cross-sectional, or synthetic cohort, measure based on what is known as the ***reduced events*** approach to demographic analysis. This approach was developed by the French demographer Louis Henry, and is described by Wunsch and Termote (1978: Chapters 1 and 2). It is applicable to demographic processes (such as fertility, first marriage and divorce) which ‘do not exclude members of [a] cohort from [future] observation’ (as mortality and emigration do – they cause individuals to disappear from a population) (Wunsch and Termote 1978: 45). At the core of the reduced events approach are annual ratios of events at each single-year duration d since ***‘event-origin’*** (the event initiating exposure to risk – attainment of the minimum legal age for marriage in a study of first marriage) to the population at duration d who were at risk ***at that earlier life cycle stage corresponding to event-origin*** (the ***total*** population at duration d in a study of the process of first marriage, since ***everyone*** surviving at that duration was at risk of marrying for the first time when they reached the minimum legal age for marriage).

In an analysis of the first marriage process, which because its tempo typically differs for women and men (women generally marry younger) should be refined by sex, these ratios are ratios of first marriages of females or males aged x in year y (symbolized by F(s,x), where s stands for ‘sex’) to the size of the mid-year female or male population aged x in year y (P(s,x)). In case you are puzzled by reference here to ages, x, instead of durations, d, the former are simply another way of expressing the latter. Each duration, d, corresponds to an age, x, such that x is the sum of d and the legal minimum age for marriage. Summing the ratios (reduced events) just defined over all ages, x, commencing at the minimum legal age for marriage is the same as summing them over all durations, d, beyond that legal minimum age, and yields the TFMR for sex s in year y. Thus the TFMRs for females and males are given by:

$$\mathrm{TFMR_f} = \Sigma_{x=l,\omega}\ (\mathrm{F}(\mathrm{f},x)/\mathrm{P}(\mathrm{f},x))\,1{,}000 \tag{5.5}$$

$$\mathrm{TFMR_m} = \Sigma_{x=l,\omega}\ (\mathrm{F}(\mathrm{m},x)/\mathrm{P}(\mathrm{m},x))\,1{,}000 \tag{5.6}$$

Where $\mathrm{TFMR_f}$ and $\mathrm{TFMR_m}$ are the female and male total first marriage rates; F(f,x) and F(m,x) = first marriages of females aged x and males aged x during year y; P(f,x) and P(m,x) = the mid-year female and male populations aged x in year y; l = the legal minimum age for marriage for the relevant sex; ω = the oldest age at which any first marriage is recorded by the relevant sex in year y.

The TFMR indicates ***the proportion of a female or male birth cohort who would EVER marry IF that cohort experienced the age-specific first marriage ratios prevailing among women/men in the year for which it was calculated.*** In other words, as already intimated, it is a synthetic cohort measure; a period, or cross-sectional, measure for which there is a cohort type of interpretation. You can probably appreciate the analogy to the total fertility rate as well. The TFR measures the average number of children a female birth cohort would have ***if*** it experienced through life the age-specific fertility rates prevailing in the year for which it was calculated.

You should recognize that ***the TFMR can take on theoretically impossible values***; values greater than 1,000, which imply greater than universal marriage. This phenomenon aroused considerable debate when first noted some years ago, but is no cause for alarm. It merely indicates that one or more of the forces that were seen in Chap. 3 to cause period measures to take on exaggerated values compared to equivalent cohort measures is at work. ***Either*** some historical circumstance has caused an abnormally large number of first marriages to take place during the year in question ***or*** a trend to earlier marriage is in progress, ***or*** both. A ***heaping*** of first marriages has occurred cross-sectionally, much the same as a heaping of births may occur cross-sectionally and result in an exaggerated peak in the trend of the total fertility rate compared to that in the trend of the cohort completed fertility rate (refer Chap. 3, Fig. 3.8). First marriage has been such a widespread, or near-universal, experience over long periods of demographic history that heaping of this kind easily can produce the 'theoretically impossible' result described. Its occurrence, however, merely implies that a compensating ***downturn*** in annual (cross-sectional) first marriages occurred earlier or is to be expected in future. Moreover, it merely reflects a reality that there is no good reason to suppress – the reality that first marriages ***can*** heap in individual years at levels unsustainable in the longer term. Just as trends in the TFMR can peak at levels that are artificially high, so, too, they can trough at levels decidedly lower than any real cohort will ever experience. Instead of first marriages heaping cross-sectionally in response to historical events favourable to marriage and/or a trend to earlier marriage, substantial cross-sectional ***deficits*** can occur as a result of ***un***favourable historical events and/or a trend to later marriage.

It is possible to compute ***real cohort equivalents*** of the TFMR, based on first marriage events distributed diagonally through a Lexis diagram rather than vertically up its columns. However, because we generally can't afford to wait around while a birth cohort passes through all possible marriageable ages, the tendency is to cumulate age-specific first marriage ratios to some upper limit which is beyond the ***main*** marrying ages, but well below the ***oldest*** age at which first marriages occur. Choice of this upper limit is arbitrary, but the aim is to capture the bulk of first marriage activity and ignore only ages at which few first marriages occur. Putting it another way, the aim is to facilitate production of a trend line whose ***shape*** is reliable, even though it may be ***displaced slightly downward*** from its position were it possible to cumulate over ***all*** marriageable ages. Suitable upper limits when dealing with a population like Australia's might be exact ages 35 for females and 40 for males, although with more recent trends to later marriage a case might be made for exact ages 40 and 45 respectively.

When we cumulate age-specific first marriage ratios for real cohorts of women or men, starting at the youngest age at which marriages occur and working up, the measure we obtain is a cohort ***cumulative first marriage rate*** (CFMR) to whatever upper age limit we have specified. This tells us the proportion of the cohort who had married by the time it reached that age limit, and because we are now dealing with the experience of a ***real*** cohort we should not obtain results which imply more than universal marriage. What was only a ***theoretical*** impossibility in the context of the synthetic cohort measure (TFMR) is an ***actual*** impossibility in the context of the equivalent real cohort measure (CFMR). The relevant equations are:

$$\mathrm{CFMR_{f,u}} = \Sigma_{\mathrm{x=l,u-1}} \left(\mathrm{F}\,(\mathrm{f,x})\,/\mathrm{P}\,(\mathrm{f,x})\right) 1{,}000 \tag{5.7}$$

$$\mathrm{CFMR_{m,u}} = \Sigma_{\mathrm{x=l,u-1}} \left(\mathrm{F}\,(\mathrm{m,x})\,/\mathrm{P}\,(\mathrm{m,x})\right) 1{,}000 \tag{5.8}$$

Where $\mathrm{CFMR_{f,u}}$ and $\mathrm{CFMR_{m,u}}$ are the female and male cohort cumulative first marriage rates to exact age u; F(f,x) and F(m,x) = first marriages at age x of females or males who were members of the relevant birth cohort; P(f,x) and P(m,x) = the size of the female or male birth cohort at the one point in time when all its members were aged x; l = the legal minimum age for marriage for the relevant sex; u = the exact age defining the upper limit to which age-specific first marriage ratios are being cumulated.

By way of example you may like to refer back to Fig. 3.11 in Chap. 3. This shows for twentieth century Australian female birth cohorts CFMRs to exact age 35. The equivalent graph for males is presented as Fig. 5.1. Both graphs use first marriage data up until 2002, beyond which the necessary single-year-of-age data ceased being published. The prevalence of female marriage by exact age 35 is shown to have risen from about 82 % for the 1899–1900 birth cohort to well over 95 % for cohorts born in the early and mid-1930s, before falling sharply again, especially among cohorts born in the 1950s. The prevalence of male marriage by exact age 40 rose from a similar level (82 %) for the 1899–1900 birth cohort to a lower peak (a little over 90 %), before also latterly declining again. In other words, marriage became progressively more universal among birth cohorts of the first thirty or so years of the twentieth century, especially female birth cohorts, but became less universal again among cohorts born following the Second World War.

You will note in both Fig. 3.11 and Fig. 5.1 that the processes of cumulating age-specific first marriage ratios have been carried out in stages, yielding intermediate plots of cumulative first marriage rates to various exact ages below the respective upper limits of 35 and 40 years. These can be useful adjuncts to plots of CFMRs as defined by Eqs. 5.7 and 5.8. For example, it is interesting to compare, between the 1:1:46 and 1:1:73 time lines, female plots to exact ages 20 and 22 in Fig. 3.11 with the male plots to exact ages 22 and 25 in Fig. 5.1. The former rise steeply then plateau while the latter rise more steadily throughout the 27-year period. Exact ages 20 and 22 for females, and 22 and 25 for males identify roughly equivalent stages in the marriage process, and this differential pattern of change reflects a steeper, quicker decline in ages at first marriage among females in post-war Australia in contrast to a more gradual decline among males. A major factor in this difference was a marked gender imbalance in the ***marriage market*** through the 1950s. Males at this time faced what demographers term a ***marriage squeeze***. Partly

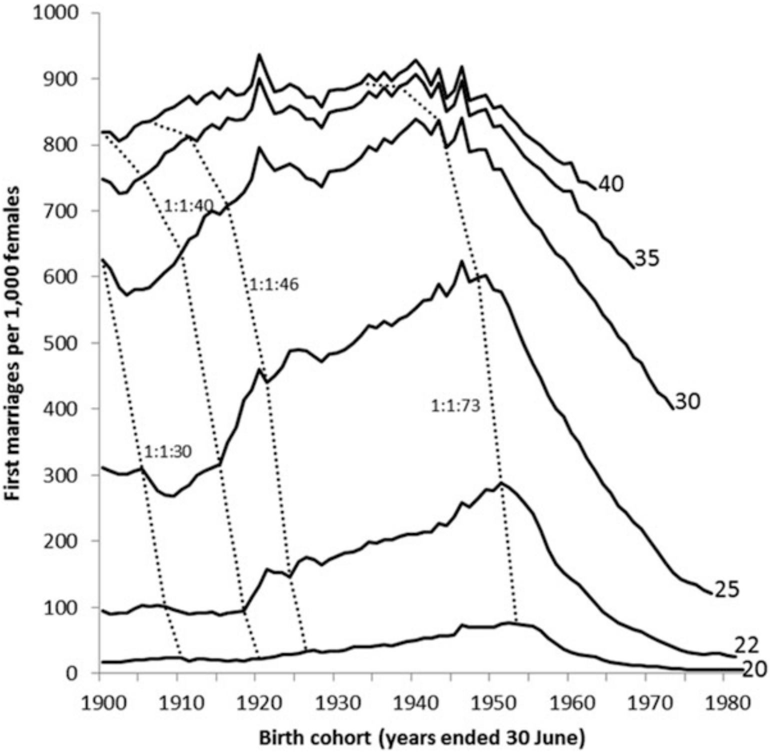


Fig. 5.1 Cumulative first marriage rates to selected exact ages for Australian male birth cohorts (Source: Adapted from Carmichael (1988: Figure 5))

because of depressed fertility during the early 1930s but mainly because of heavy male dominance of the large volume of early post-war immigration to Australia, marriageable men faced a shortage of potential marriage partners. The competition for brides this produced exerted strong pressure on women to marry younger, but left some men unable to marry as early in life as they might have wished until, with the first baby boom cohorts entering it, the marriage market improved for them during the 1960s. This phenomenon also largely explains why the male CFMR to exact age 40 peaks at a lower level than the female CFMR to exact age 35. Over a period when almost everyone wanted to marry, some men missed out. There were too few women to go around.

Measures of the Timing of the First Marriage Process

We have discussed the TFMR and CFMR as measures of the ***intensity*** of the process of first marriage for synthetic and real birth cohorts respectively. The other dimension of the first marriage process in which demographers are interested is its tempo, or the ***timing*** of first marriage. They want to know, in general terms, whether ages at first marriage are relatively young or relatively old, and whether they are rising, falling or remaining constant over time.

A commonly used ***period*** measure of marriage timing is the ***median age at first marriage***. The median is a preferable measure of the ‘average’ age at which first marriages occur to the ***mean*** age at first marriage, because distributions of first marriages by age tend to be heavily skewed to the left. That is, if the number of first marriages by age is plotted, the resulting graph generally has a pronounced peak towards the lower end of the age range over which first marriages occur, and a long ‘tail’ extending over the middle and upper end of that range. The few first marriages that take place at more advanced ages have an inordinate influence on the mean age at first marriage, tending to push it above the age at which first marriage is in fact most common. The median is not subject to this bias, is likely to coincide better with the graphical peak of the distribution of first marriages by age, and is thus a more reliable measure of ***central tendency***, or ‘average’ behaviour.

As with any median, to obtain the median age at first marriage one simply takes an annual distribution of first marriages by age and determines the age above and below which exactly half of first marriages occurred. It should be calculated separately for females and males, and where it falls within a single-year age group rather than neatly at the boundary between two successive age groups (as it almost always will) its precise value is obtained by linear interpolation within that age group. The relevant equation is:

$$\mathrm{Md} = \mathrm{X} + (\mathrm{T}/2 - \Sigma_{\mathrm{i=l,x-1}}\ \mathrm{M_i})\ /\mathrm{M_x} \tag{5.9}$$

Where T = the total number of first marriages in year y; $\mathrm{M_i}$ = first marriages at age i in year y; l = the lowest, or youngest, age at which first marriages occurred in year y; X = the single-year age group within which the median age at first marriage lies (i.e., the youngest age group for which $\sum_{\mathrm{i=l,\,x}} \mathrm{M_i}$ equals or exceeds T/2).

The median age at first marriage is in common use. It has a limitation, though, in not taking account of the varying sizes of single-year age groups. Other things being equal, the larger the age group the more first marriages it will produce. Trends over time in the median age at first marriage can thus be affected by shifting age structures, especially over the ages at which first marriage is most common. A better procedure for obtaining period trends in marriage timing, and one which controls for this source of change, is to ***divide the total first marriage rate into quartiles and ascertain the ages at which each quartile is reached***. To do this:

1. Find the values of quarter, half and three-quarters of the TFMR. These correspond with the first, second and third quartiles of the age distribution of first marriages respectively.
2. Begin cumulating the F(f,x) / P(f,x) or F(m,x) / P(m,x) ratios from the equation for the TFMR (Eq. 5.5 or Eq. 5.6 as appropriate) from x = l, the lowest or

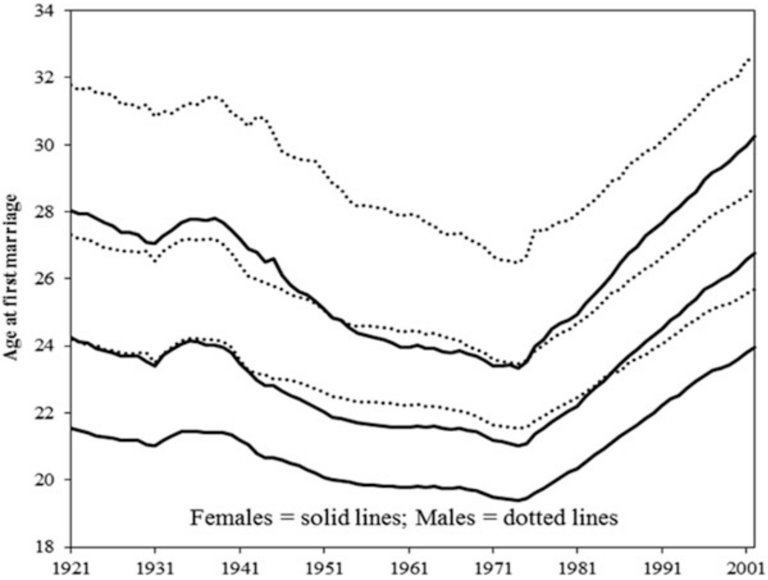


Fig. 5.2 Quartiles of the age distributions of first marriages for Australian females and males married between 1921 and 2002 (Source: Adapted from Carmichael (1988: Figure 3))

youngest age at which first marriages occur, on upwards. In doing so, identify the exact ages at which the three quartiles are reached.

3. Usually you will find that by cumulating to a value of $x = m$ you don't quite reach a quartile, whereas by cumulating to $x = m + 1$ you have gone past it. This means that the quartile occurs at an ***exact*** age that is greater than $m + 1$, but less than $m + 2$ years (x, remember, is a single-year age group, or an age in ***completed*** years). To find the exact age corresponding to a quartile, use:

$$QA = m + 1 + [Q - \Sigma_{x=l,m} (F(s,x) / P(s,x))] / (F(s, m+1) / P(s, m+1)) \tag{5.10}$$

Where QA stands for 'quartile age'; Q = the relevant quartile value of the TFMR (i.e., quarter, half or three-quarters of the TFMR, depending on whether one is evaluating the first, second or third quartile of the age distribution of first marriages); l = the youngest age at which first marriages occur; m = the ***highest*** value of x (age) for which the TFMR summation on the righthand side of Eq. 5.5 or Eq. 5.6 lies ***below*** Q; F(s,x) and P(s,x) have the same meanings as in Eqs. 5.5 and 5.6, with s denoting 'sex' (either f (female) or m (male)).

Quartile ages at first marriage for Australian females and males marrying between 1921 and 2002 are shown in Fig. 5.2. Single-year-of-age first marriage data required to calculate them have not been published since 2002. Naturally the lowest trend line for each sex plots the age by which 25 % of first marriages occurred; the middle line the age by which half occurred; and the highest line the age by which 75 % occurred. Ages at first marriage declined through the 1920s, rose again from the Great Depression to plateaux in the middle and late 1930s, then followed protracted downward courses from the outbreak of World War 2 until the early 1970s. Declines during the War itself were steep as couples shortly to be separated when young men left for war service took a 'now or never' attitude to marriage and because of their imminent separation disregarded its usual economic prerequisites. There were also 12–15 thousand marriages involving visiting American servicemen during the War.

Following the War continued downward trends were initially steeper for females, reflecting pressure to marry exerted by the previously noted excess of marriageable males in the population. Female ages at first marriage in particular looked to have stabilized during the early and mid-1960s, but the late 1960s and early 1970s brought further declines as oral contraception permitted very early marriage on the basis that parenthood could confidently be deferred thereafter. At this time peer pressure on the young to be sexually active had become intense, but many parents still preached premarital chastity, and early marriage coupled with reliable contraception was for some an ideal way of resolving these conflicting pressures.

After the early 1970s the trend to earlier marriage dramatically reversed, and by the mid-1990s most quartile ages at first marriage had regained and surpassed their 1921 levels. From the pattern of very early marriage followed by a childless, dual income start to married life that the pill fostered in the late 1960s and early 1970s it was a short step to couples simply cohabiting without getting married, especially with a rising divorce rate recommending a more cautious approach to marriage. The growing acceptability of consensual unions has been one factor behind the trend to later marriage, but both trends have also been products of the change feminism induced in young women's priorities in late adolescence and early adulthood, and of increased emphasis by the young of both sexes on maintaining their individual autonomy. Feminism saw education, establishing careers, travel and having a good time assume higher priorities and finding a husband a lower priority; the quest for more autonomy overlapped with feminism, but also made young men reluctant to commit to marriage. A rapid decline in the incidence of 'shotgun' marriage over the period 1971–1976, largely due to freer access to abortion and aided by increased consensual partnering (hence more reliable contraception) and the introduction of a welfare benefit for single mothers, was also something of a trigger to the new trend. It eliminated many marriages at the very youngest ages in which brides had been especially likely to be already pregnant, and saw the Judaeo-Christian view that marriage was the honourable response to unplanned premarital pregnancy widely rejected almost overnight.

You should note that the measures plotted in Fig. 5.2 measure marriage timing ***among those who did marry***. Beware of measures which are ***claimed*** to be, for example, median ages at first marriage, but which actually indicate the age by which 50 % of ***all people*** (not 50 % of ***those who married***) married. There is sometimes a particular temptation to present this type of measure when examining the first marriage experience of ***real*** birth cohorts, because part of their first marriage experience still lies in the future and this makes it impossible to calculate a genuine median. If ***everyone*** marries, the age by which 50 % have married is indeed the median (the halfway point in the age distribution of first marriages), but if, say, only 80 % of a cohort marries, it is ***five-eighths*** of the way up the age distribution of first marriages. Thus this measure is not a 'pure' measure of marriage timing; its meaning depends on the ***intensity*** of the first marriage process and, critically, ***changes as intensity changes***. Where only 80 % of a cohort ever married the median age at first marriage would be the age by which 40 % married (i.e., half of those who ultimately ever married). Failure to marry at all is a different phenomenon from the timing of marriage among those who do marry.

The procedure for obtaining cross-sectional quartiles of the age distribution of first marriages summarized in Eq. 5.10 can be extended to studying the first marriage experience of real birth cohorts by focusing on first marriages of those marrying by some arbitrarily selected exact age u (standing for 'upper limit'). Proceed as follows:

1. Find the values of quarter, half and three-quarters of the cohort CFMR to exact age u. These correspond with the first, second and third quartiles of the age distribution of first marriages taking place before exact age u respectively.
2. Begin cumulating the F(f,x) / P(f,x) or F(m,x) / P(m,x) ratios from the equation for the CFMR (Eq. 5.7 or Eq. 5.8 as appropriate) from x = l, the lowest or youngest age at which first marriages occur, on upwards. In doing so, identify the exact ages at which the three quartiles are reached.
3. Do this by again applying Eq. 5.10, with obvious modifications to the meaning of terms in that equation. Q = the relevant quartile value of the cohort ***CFMR to exact age u***; m = the highest value of x (age) for which the ***CFMR*** summation on the righthand side of Eq. 5.7 or Eq. 5.8 lies below Q; F(s,x) and P(s,x) have the same meanings as in Eqs. 5.7 and 5.8, with s denoting 'sex' (either f (female) or m (male)).

Nuptiality Tables

The life table principles discussed in relation to the study of mortality in Chap. 4 may also be applied to the study of first marriage. Consider a cohort of never married persons of one sex attaining the minimum age for marriage, and then passing through the marriageable ages. Individuals are removed from that population at risk of first marriage by one of two processes – either they marry, or they die without having married. On experiencing either of those events a person ceases to be at risk of marrying for the first time.

If we had all of the information about marriages and deaths occurring to our cohort we could calculate numbers ‘surviving’ never married at each birthday and probabilities of marrying and of dying without having married between successive birthdays. These probabilities would enable us to construct a ***double decrement*** life (or attrition) table for the cohort; i.e., a life table in which there were ***two*** modes of depletion of the initial radix population. This table would allow us to estimate for persons surviving never married at each birthday the average number of years left ***in the never married state*** before either marrying or dying.

We have talked here of tracing the depletion through first marriage and mortality of a ***real cohort*** of never married males or females. With the conventional life table we didn’t trace the mortality of real cohorts because they take so long to die out. Rather we traced the mortality of cross-sectional ***synthetic*** cohorts. We do the same when dealing with first marriage. We construct what is called a ***net nuptiality table*** on the basis of sex-age-specific first marriage and death rates ***in a calendar year***. As with the conventional life table the key to constructing a net nuptiality table is the calculation of attrition probabilities – probabilities of first marriage (v_x) and of death (q_x). Before progressing further with net nuptiality tables, however, there is a simpler type of nuptiality table to consider – the ***gross nuptiality table***.

Gross Nuptiality Tables

A gross nuptiality table ***ignores death*** as a mode of attrition from the radix of never married persons. It is therefore a ***single*** decrement table directly analagous to the standard life table. The rationale for ignoring mortality is that attrition from this source is generally so low compared to attrition due to marriage over the main marrying ages that it is inconsequential. Generating a gross nuptiality table requires probabilities of first marriage (v_x) to be obtained, with remaining columns of the table then derived exactly as they would be in a standard single-year-of-age life table (writing v_x for q_x in the relevant equations).

The literature provides several equations for obtaining v_x values. Recalling the equations for q_x in terms of the age-specific death rate M_x which were derived for the standard life table in Chap. 4 (Eqs. 4.7 and 4.8), the analagous equations for v_x in terms of age-specific first marriage rates n_x are given by:

$$v_x = n_x/\left(1 + {}^1\!/_2\, n_x\right) = 2n_x/(2 + n_x) \qquad (5.11)$$

Where n_x is first marriages of females or males aged x in year y divided by the mid-year population of ***never married*** females or males aged x in year y.

Equation 5.11 is for use when marriage registration data are available to provide numerators for n_x calculations. There is also a procedure available which allows v_x values to be estimated from census data giving sex-age-specific proportions never married. It is a relatively crude approach, and should only be used when no other (registration-based) option is open. The relevant equation is:

$$v_x = {}^1\!/_2\,[((s_{x-1} - s_x)/s_{x-1}) + ((s_x - s_{x+1})/s_x)] \tag{5.12}$$

Where s_x = the proportion of females or males never married at age x.

An example of a gross nuptiality table is shown as Table 5.1. Several features of this table should be noted. First, apart from age 0 the first age shown is age 14, the youngest age at which any marriage occurs in the population. There is no point in showing intermediate ages because no first marriage activity is taking place, and the table assumes zero mortality. Note, though, that ${}_{14}L_0$ has a value equal to 14 times the radix (1,400,000), since everybody is assumed to survive the first 14 years of their lives never married, and therefore to contribute 14 person-years of exposure to risk. This exposure must be taken into account in calculating $e^o{}_0$, which is perhaps the most important number in a gross nuptiality table (its interpretation is discussed below).

Second, Table 5.1 is restricted to ages below 50, so that the T_x column gives total person-years lived in the never married state between exact age x ***and exact age 50*** (not at ***all*** ages beyond exact age x) and $e^o{}_x$ gives the average number of years remaining in the never married state ***before attaining exact age 50*** for persons still never married at exact age x. Why, you might ask, do we restrict the range of ages covered by a gross nuptiality table in this way? One reason is that our single mode of attrition, first marriage, is not a universal experience as death is, so that our radix of 100,000 never married persons ***never completely 'dies out'*** (in the sense of all of them getting married). Eventually a point (an age) is reached at which the great bulk of the first marriage activity that will ever take place ***has*** taken place, and there is little to be gained from considering what happens at older ages. A second reason for truncating at age 50 is that the basis upon which we ignore mortality in generating a gross nuptiality table – that it is an inconsequential mode of decrement compared to first marriage – is only sustainable over ages where most first marriages occur. It certainly is not sustainable at older ages, where death takes over as the main force of attrition among the few remaining survivors never married. Thus we truncate before reaching the point at which our having disregarded mortality becomes a problem.

A third point that may have occurred to you is that, while mortality may generally be very low through the main marrying ages, it may not be so low at ages ***below*** the minimum age for marriage, and especially in infancy. The gross nuptiality table assumes zero mortality at these ages as well, and this influences the value of $e^o{}_0$, which as already mentioned is a particularly crucial number in a gross nuptiality table. The quantity $e^o{}_0$ gives the average number of years spent in the never married state before age 50 ***on the assumption that a person remains alive for at least***

Table 5.1 Gross nuptiality table for Australian women, 1980

Age (x)	Never married females	Female first marriages	v_x	l_x	${}_nL_x$	T_x	$e^o{}_x$
0				100, 000	1, 400, 000	2, 719, 169	27.2
14	119,633	4	0.000033	100, 000	99, 998	1, 319, 169	13.2
15	121,024	31	0.000256	99, 997	99, 984	1, 219, 171	12.2
16	125,445	787	0.006254	99, 971	99, 658	1, 119, 187	11.2
17	125,266	2144	0.016970	99, 346	98, 503	1, 019, 529	10.3
18	122,915	6379	0.050585	97, 660	95, 190	921, 026	9.4
19	113,468	10,102	0.085235	92, 720	88, 768	825, 836	8.9
20	95,051	12,081	0.119505	84, 817	79, 749	737, 068	8.7
21	80,574	11,962	0.138201	74, 681	69, 520	657, 319	8.8
22	63,022	10,111	0.148521	64, 360	59, 580	587, 799	9.1
23	49,906	7480	0.139432	54, 801	50, 980	528, 219	9.6
24	40,798	5664	0.129818	47, 160	44, 099	477, 238	10.1
25	30,914	4214	0.127615	41, 038	38, 419	433, 140	10.6
26	24,029	3037	0.118876	35, 801	33, 673	394, 720	11.0
27	20,025	2311	0.109109	31, 545	29, 824	361, 048	11.4
28	16,873	1825	0.102611	28, 103	26, 661	331, 224	11.8
29	13,783	1337	0.092516	25, 219	24, 053	304, 563	12.1
30	11,555	1021	0.084621	22, 886	21, 918	280, 510	12.3
31	9518	822	0.082787	20, 949	20, 082	258, 593	12.3
32	8570	638	0.071774	19, 215	18, 525	238, 510	12.4
33	7883	503	0.061835	17, 836	17, 284	219, 985	12.3
34	6111	413	0.065373	16, 733	16, 186	202, 701	12.1
35	6090	297	0.047607	15, 639	15, 267	186, 515	11.9
36	5790	224	0.037953	14, 895	14, 612	171, 248	11.5
37	4423	180	0.039884	14, 329	14, 043	156, 636	10.9
38	4389	161	0.036021	13, 758	13, 510	142, 592	10.4
39	3985	116	0.028691	13, 262	13, 072	129, 082	9.7
40	3737	109	0.028748	12, 882	12, 696	116, 010	9.0
41	3578	100	0.027563	12, 511	12, 339	103, 314	8.3
42	3436	74	0.021307	12, 166	12, 037	90, 975	7.5
43	3570	64	0.017767	11, 907	11, 801	78, 938	6.6
44	3193	55	0.017078	11, 696	11, 596	67, 137	5.7
45	3017	49	0.016110	11, 496	11, 403	55, 541	4.8
46	2991	34	0.011303	11, 311	11, 247	44, 138	3.9
47	3050	42	0.013676	11, 183	11, 106	32, 891	2.9
48	2759	35	0.012605	11, 030	10, 960	21, 785	2.0
49	3082	38	0.012254	10, 891	10, 824	10, 824	1.0
50	3135	38	0.012048	10, 757			

50 years. It is a useful index of the urgency with which a population marries, and changes over time in its level can be plotted. It is important to realise, however, that such changes can be produced ***either*** by changes in marriage timing ***or*** by changes in the propensity to get married at all. The larger the proportion of the radix population that reaches age 50 still never married the higher e^o_0 will be, other things (i.e., the timing of marriage among those who do marry) remaining equal. Thus e^o_0 is ***not***, strictly speaking, a measure of marriage timing akin to the median age at first marriage.

The l_x column of a gross nuptiality table is of particular interest. Dividing l_x by the radix l_0 tells us the proportion of a cohort experiencing the cross-sectional probabilities of first marriage underlying the table who would remain never married at exact age x. Subtracting from 1.0 then gives us the proportion ***ever*** married by exact age x, and proportions of this type can be plotted for successive synthetic cohorts (each the subject of a separate gross nuptiality table) to establish trends over time.

Net Nuptiality Tables

We can obtain the two sets of probabilities q_x and v_x needed for the construction of a net nuptiality table from age-specific death rates, M_x, and age-specific first marriage rates, n_x, using equations already given. To find q_x values use equation 4.7 or equation 4.8 from Chap. 4; to find v_x values use Eq. 5.11 above. Strictly speaking these equations are slightly imprecise vehicles for obtaining q_x and v_x values for use in a double decrement context. But they are satisfactory, and we can get away with using them, because of the overwhelming importance of first marriage, and the minimal importance of death, as modes of attrition from the never married population through the main marrying ages. More refined procedures are available, but little is gained from using them in constructing net nuptiality tables. Note, though, that this is not the case with ***all*** double decrement life tables. For other types of double decrement tables in which ***both*** modes of attrition are reasonably important it is not acceptable to obtain attrition probabilities using equations modelled directly on Eqs. 4.7 and 4.8 from Chap. 4 and Eq. 5.11 above; a more refined procedure ***must*** be used. One way of thinking of our ability to get away with a relatively simple approach in the case of net nuptiality tables is to note that, because of the minimal importance of death, we are very close to really only having ***one*** mode of attrition. That, of course, is also the rationale for often making do with a ***gross*** nuptiality table.

Another point to note is that, strictly speaking, the mortality probabilities in a net nuptiality table should pertain to never married persons ***only***, not the total population. Once again, though, because mortality is minimal over the prime marrying ages there is nothing of consequence to be gained by being rigorous in this matter, and total population probabilities, which are likely to be readily available from a standard life table, generally are used.

A net nuptiality table for White females in the United States in the late 1950s is presented as Table 5.2. Its first two columns give the probabilities of first marriage and of dying at age 0 and at all ages from 14 to 65 and over. The minimal importance of mortality from the late teens to the early thirties, when first marriage is most prevalent, is clear. Columns 4 and 5, which give values of d'_x and f'_x, the numbers of deaths and first marriages, respectively, at age x, reinforce this. These are obtained from the equations:

$$d'_x = q_x\,(1 - v_x/2)\,l'_x \tag{5.13}$$

$$f'_x = v_x\,(1 - q_x/2)\,l'_x \tag{5.14}$$

The l'_x column, column 3 of Table 5.2, is the net nuptiality table equivalent of the l_x column of the conventional life table; it gives the number of members of the radix population (i) still alive and (ii) still never married at exact age x. It is obtained using the equation:

$$l'_x = l'_{x-1} - d'_{x-1} - f'_{x-1} \tag{5.15}$$

It may seem strange that Eqs. 5.13 and 5.14 require a value of l'_x to obtain values of d'_x and f'_x, and Eq. 5.15 requires values of d'_{x-1} and f'_{x-1} to obtain values of l'_x. 'Where do I start?' you might ask. The answer is, by evaluating Eqs. 5.13 and 5.14 for $x = 0$, since l'_0 is known (the nuptiality table radix = 100,000). It is then possible to obtain l'_1 using Eq. 5.15, then d'_1 and f'_1 using Eqs. 5.13 and 5.14, and so on. We simply shuffle between Eqs. 5.13/5.14 and 5.15, working our way down the net nuptiality table from $x = 0$, increasing x by 1 each time.

Two idiosyncrasies of Table 5.2 should be noted. First, entries for values of x in the range 1–13 are excluded because no first marriages take place at those ages. The same is true at age 0, but this age is represented because (i) we need to know l'_0, the radix of the table, and (ii) $e'^{o}{}_x$, the average number of years of single life remaining ***at birth***, is an important feature of a net nuptiality table. Deaths at age 0 $\left(d'_0 = 1,964\right)$ are shown, and any attrition between exact ages 1 and 14 is also due to mortality. We can ascertain its extent from:

$${}_{13}d'_1 = l'_1 - l'_{14} = \left(l'_0 - d'_0\right) - l'_{14} = 100{,}000 - 1{,}964 - 97{,}411 = 625$$

The second idiosyncrasy pertains to the bottom of the nuptiality table where the terminal age group is 65 and over. Here Eqs. 5.13 and 5.14 are not used to obtain d'_x and f'_x (which are more correctly specified as d'_{x+} and f'_{x+}). The sum of the probabilities of first marriage (v_{65+}) and death while still never married (q_{65+}) is 1.00000 (since everyone who doesn't eventually marry ultimately dies never

Table 5.2 Net nuptiality table for white females in the United States, 1958–1960

	v_x	q_x	l'_x	d'_x	f'_x	N'_x	$\%N'_x$	L'_x	T'_x	e'^o_x
Age (x)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	0.00000	0.01964	100,000	1,964	0	94,937	94.9	98,319	2,157,010	21.57
14	0.01167	0.00036	97,411	35	1,137	94,937	97.5	96,948	789,410	8.10
15	0.03072	0.00041	96,239	39	2,956	93,800	97.5	94,942	692,462	7.20
16	0.06361	0.00047	93,244	42	5,930	90,844	97.4	90,532	597,520	6.41
17	0.10919	0.00051	87,272	42	9,627	84,914	97.3	82,839	506,988	5.81
18	0.18499	0.00054	77,703	38	14,370	75,387	97.0	70,660	424,149	5.46
19	0.21168	0.00055	63,295	31	13,395	61,017	96.4	56,466	353,489	5.58
20	0.23286	0.00056	49,869	25	11,609	47,622	95.5	43,907	297,023	5.96
21	0.25988	0.00058	38,235	19	9,934	36,013	94.2	33,093	253,116	6.62
22	0.27009	0.00060	28,282	15	7,636	26,079	92.2	24,250	220,023	7.78
23	0.24127	0.00062	20,631	11	4,976	18,443	89.4	17,955	195,773	9.49
24	0.20919	0.00063	15,644	9	3,272	13,467	86.1	13,894	177,818	11.37
25	0.19091	0.00065	12,363	7	2,359	10,195	82.5	11,112	163,924	13.26
26	0.16344	0.00068	9,997	6	1,633	7,836	78.4	9,126	152,812	15.29
27	0.13507	0.00071	8,358	6	1,129	6,203	74.2	7,763	143,686	17.19
28	0.13374	0.00074	7,223	5	966	5,074	70.2	6,721	135,923	18.82
29	0.11805	0.00079	6,252	5	738	4,108	65.7	5,862	129,202	20.67
30	0.09403	0.00085	5,509	4	518	3,370	61.2	5,235	123,340	22.39
31	0.08532	0.00091	4,987	4	425	2,852	57.2	4,765	118,105	23.68
32	0.07274	0.00097	4,558	4	331	2,427	53.2	4,385	113,340	24.87
33	0.06827	0.00105	4,223	4	288	2,096	49.6	4,072	108,955	25.80
34	0.05659	0.00113	3,931	4	222	1,808	46.0	3,814	104,883	26.68
35	0.04851	0.00122	3,705	4	180	1,586	42.8	3,610	101,069	27.28
36	0.04359	0.00133	3,521	5	153	1,406	39.9	3,440	97,459	27.68
37	0.04047	0.00145	3,363	5	136	1,253	37.3	3,291	94,019	27.96
38	0.03740	0.00158	3,222	5	120	1,117	34.7	3,158	90,728	28.16
39	0.03445	0.00174	3,097	5	107	997	32.2	3,040	87,570	28.28
40	0.03166	0.00190	2,985	6	94	890	29.8	2,934	84,530	28.32
41	0.02902	0.00209	2,885	6	84	796	27.6	2,839	81,596	28.28
42	0.02660	0.00229	2,795	6	74	712	25.5	2,754	78,757	28.18
43	0.02438	0.00252	2,715	7	66	638	23.5	2,678	76,003	27.99
44	0.02237	0.00276	2,642	7	59	572	21.6	2,608	73,325	27.75
45	0.02028	0.00303	2,576	8	52	513	19.9	2,546	70,717	27.45
46	0.01886	0.00331	2,516	8	47	461	18.3	2,488	68,171	27.09
47	0.01768	0.00362	2,461	9	43	414	16.8	2,435	65,683	26.69
48	0.01648	0.00396	2,409	9	40	371	15.4	2,384	63,248	26.25
49	0.01530	0.00432	2,360	10	36	331	14.0	2,337	60,864	25.79
50	0.01415	0.00473	2,314	11	33	295	12.7	2,292	58,527	25.29
51	0.01300	0.00517	2,270	12	29	262	11.5	2,249	56,235	24.77
52	0.01189	0.00560	2,229	12	26	233	10.5	2,210	53,986	24.22
53	0.01085	0.00601	2,191	13	24	207	9.4	2,172	51,776	23.63

(continued)

Table 5.2 (continued)

	v_x	q_x	l'_x	d'_x	f'_x	N'_x	$\%N'_x$	L'_x	T'_x	e'^o_x
Age (x)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
54	0.00984	0.00642	2,154	14	21	183	8.5	2,136	49,604	23.03
55	0.00888	0.00687	2,119	14	19	162	7.6	2,102	47,468	22.40
56	0.00799	0.00740	2,086	15	17	143	6.9	2,070	45,366	21.75
57	0.00718	0.00805	2,054	16	15	126	6.1	2,038	43,296	21.08
58	0.00650	0.00886	2,023	18	13	111	5.5	2,008	41,258	20.39
59	0.00590	0.00981	1,992	19	12	98	4.9	1,977	39,250	19.70
60	0.00540	0.01088	1,961	21	11	86	4.4	1,945	37,273	19.01
61	0.00496	0.01203	1,929	23	10	75	3.9	1,913	35,328	18.31
62	0.00458	0.01325	1,896	25	9	65	3.4	1,879	33,415	17.62
63	0.00432	0.01454	1,862	27	8	56	3.0	1,845	31,536	16.94
64	0.00413	0.01592	1,827	29	7	48	2.6	1,809	29,691	16.25
65+	0.02289	0.97711	1,791	1,750	41	41	2.3	27,882	27,882	15.57

Source: Adapted from Shryock, Siegel and Associates (1973: 561)

married), and d'_{65+} and f'_{65+} are found by simply apportioning l'_{65} according to the probabilities (proportions) q_{65+} and v_{65+}.

Column 6 of Table 5.2 gives the number of persons from the radix population who marry for the first time at age x and all older ages. Denoted by N'_x it has no counterpart in a standard life table, and is calculated from:

$$N'_x = \Sigma_{i=x,65+} f'_i \tag{5.16}$$

Column 7, labelled $\% N'_x$, indicates the percentage of persons 'surviving', alive and never married, at exact age x who subsequently marry (when aged x or older). It follows that the difference between this value and 100 gives the percentage of persons surviving unmarried at exact age x who subsequently die without ever marrying. The relevant calculating equation is:

$$\%N'_x = \left(N'_x / l'_x\right) 100 \tag{5.17}$$

In Table 5.2, White females had a 94.9 % chance of ever marrying at birth, and a 5.1 % chance of dying without ever marrying. If they remained alive and unmarried at exact age 30, for example, they had a 61.2 % chance of ever marrying subsequently, and a 38.8 % chance of dying as a spinster.

Column 8 of Table 5.2 is the equivalent of the L_x column of the standard life table. It indicates the number of person-years spent alive and never married at age x (i.e., between exact ages x and x + 1) by the nuptiality table population. Values of L'_x are calculated using:

$$L'_x = {}^1\!/_2 \left(l'_x + l'_{x+1}\right) + {}^1\!/_{24} \left(d'_{x+1} + f'_{x+1} - d'_{x-1} - f'_{x-1}\right) \tag{5.18}$$

Where the expression $\frac{1}{24}(d'_{x+1} + f'_{x+1} - d'_{x-1} - f'_{x-1})$ is an adjustment for non-linearity of the survivorship function between exact ages x and x + 1.

Column 9 is the equivalent of the T_x column of the standard life table, indicating person-years lived alive and never married at all ages beyond exact age x. We have that:

$$T'_x = \Sigma_{i=x,65+} L'_i = T'_{x+1} + L'_x \tag{5.19}$$

Finally, column 10 of Table 5.2 gives the average number of years of single life remaining before marrying or dying for persons surviving alive and never married at exact age x. In a manner analagous to that used to obtain $e^o{}_x$ for a standard life table it is given by:

$$e'^o{}_x = T'_x / l'_x \tag{5.20}$$

A net nuptiality table yields various measures which can be plotted for successive synthetic cohorts (each having its own net nuptiality table). Values of l'_x at key ages can be plotted; the N'_x column can be used to divide the age distribution of first marriages into quartiles; N'_0 gives the proportion of the synthetic cohort who ever marry; $\% N'_y$, where y is the youngest age at which marriages occur, gives the proportion of persons who survive to marriageable age who ever marry; and trends in $e'^o{}_x$ for certain values of x may be of special interest. Changes in these measures may reflect changes in either first marriage or mortality levels and patterns, but once again the importance of mortality is likely to pale into insignificance compared to that of first marriage.

Two issues to conclude with. First, you may wonder whether there is any such thing as an ***abridged*** nuptiality table (net or gross). There tends not to be much point in constructing such tables because first marriage activity is heavily concentrated over a very few single years of age. Without a single-years-of-age approach over this age range a nuptiality table tells us very little. Moreover, if abridged nuptiality tables ***were*** to be constructed, very complex issues of separation, for which satisfactory ***generalized*** solutions such as exist for abridged life tables are not yet readily available, would arise.

Finally, you should be aware that while gross and net nuptiality tables are often used in the study of marriage trends and patterns they can introduce distortions into the age distribution of first marriages. We discussed earlier how to obtain quartiles of the age distribution of first marriages by finding the values of quarter, half and three-quarters of the TFMR, then determining the exact ages at which the summation of age-specific first marriage ratios which produces the TFMR reached these quartile values. We can similarly use gross and net nuptiality tables to obtain quartiles of the age distribution of first marriages, but these quartiles will rarely be identical to those yielded by the TFMR-based procedure. Why?

You will recall that the TFMR can take on ‘impossibly’ high values – values which imply that if a birth cohort were to experience the age-specific first marriage ratios recorded in the relevant year, ***more than 100 %*** of the cohort would ever marry. This phenomenon, we noted, occurred because of cross-sectional ***heaping*** of first marriages, due to circumstances being temporarily particularly conducive to marriage, a trend to earlier marriage being in progress, or both. Logically, it clearly is perfectly possible for first marriages to heap cross-sectionally at levels unsustainable in the longer term. In theory, for example, ***all*** members of a ***real*** birth cohort ***could*** marry at the minimum legal age for marriage. This is an extreme and highly improbable scenario, but if it happened ***all*** of that cohort's first marriages would take place during two consecutive calendar years during which it was passing through the one-year 'minimum age for marriage' age group. There would be huge numbers of first marriages at that age during those two years, causing the TFMR to 'go through the roof' (assuming older real cohorts were marrying 'conventionally' and contributing first marriages at older ages in the relevant two years), but at the cost of there being ***no*** first marriages contributed by the cohort at ***older*** ages in subsequent years (i.e., there would be compensating adjustments in cross-sectional data for later years).

Nuptiality tables do not recognise the possibility that first marriages might heap cross-sectionally at levels consistent with greater than universal marriage. They ***force*** the percentage ever marrying in the synthetic cohort ***never to exceed 100***. In doing so they can distort age distributions of first marriage when, for some reason, the cross-sectional incidence of first marriage is especially high. Basically the problem is that in a nuptiality table the ***number*** of first marriages a given age-specific probability of first marriage implies for the distribution of first marriages by age ***depends on probabilities at younger ages***. These determine the number of survivors (l'_x) in the nuptiality table stationary population the probability is applied to. In the extreme example cited in the previous paragraph, which would lead to a probability of first marriage of 1.0 at the minimum age for marriage in one synthetic cohort, the nuptiality table radix population for this synthetic cohort would be ***totally*** depleted after the first marriageable age (i.e. l'_{y+1} would be zero, where y was the youngest age for marriage). Probabilities of first marriage at older ages would all be applied to ***zero survivors***, producing ***zero first marriages***. Despite this example being implausible to the point of being ridiculous, the principle underlying it is valid – in a nuptiality table, probabilities of first marriage at younger ages get 'first bite of the cherry', and those at older ages are left to operate on the l'_x ***residue*** that is left. The effect, when first marriages heap cross-sectionally, can be likened to putting the lid on an overflowing rubbish bin. For the lid to fit the rubbish must be compressed to the level of the top of the bin, but as one pushes the rubbish down the degree of compaction is greater at the top of the bin, where the pressure is being applied, than it is in the middle and at the bottom. In a situation where cross-sectional first marriage activity is consistent with greater than universal marriage, nuptiality tables compress activity at older ages in order that the 'lid can be put on the bin' at 100 % ever marrying. The result is that numbers of first marriages at older ages are understated relative to numbers at younger ages, and measures like the quartiles of the age distribution of first marriages are distorted. The beauty of the TFMR-based procedure for obtaining quartiles is that the number of marriages at each age (the ratio of first marriages to total population) ***is determined independently of what has happened at younger ages***. Each one-year interval of marriageable age is treated on an identical basis, there is no artificial lid on the 'bin', and undistorted quartiles of the age distribution of first marriages result.

'What about when, instead of first marriages ***heaping*** cross-sectionally, there is a marked cross-sectional ***deficit***?' you may ask. 'Do nuptiality tables also yield a distorted picture of marriage timing in that circumstance?' The answer is 'yes'. In this situation, limited marriage activity at younger ages tends to lead to first marriage probabilities at older ages being applied to abnormally ***large*** numbers of survivors (l'_x). The consequence is that numbers of first marriages at those ages are boosted as a direct result of what is happening at younger ages, and comparatively speaking the age distribution of first marriages overemphasizes older marriageable ages. Again, the TFMR-based approach to measuring quartile ages at first marriage has the advantage that the importance of any single-year marriageable age group in no way depends on what happens at younger ages in the year to which the relevant synthetic cohort pertains.

The Singulate Mean Age at Marriage

The measures and techniques for studying the process of first marriage dealt with so far have presupposed the availability of reliable marriage registration data. In many countries, and most notably in many of those collectively described as 'less developed' or 'developing', such data are not available, and frequent use is made of the singulate mean age at marriage, or SMAM. This is a measure of marriage timing whose calculation ***requires only census or survey data showing the distribution of members of a population of a given sex by age and marital status***. More specifically, the SMAM requires as input data census or survey ***percentages of males/females never married in five-year age groups***.

The SMAM gives ***the average number of person-years lived in the single (never married) state by persons of a given sex who marry before age 50***. The following is a recipe for its calculation which explains the rationale behind each step in the calculation process. The recipe assumes one has, as a point of departure, a set of percentages of females or males never married in five-year age groups. As with other methods for analysing marriage patterns and trends, it is vital to deal separately with the two sexes.

Step 1: Assuming no marriages take place before age 15, ***add the percentages never married for ages 15–19 to 45–49***. If marriages take place before age 15, see the note after Step 8 below.

Step 2: ***Multiply this figure by 5***. This gives the number of person-years lived in the never married state between exact ages 15 and 50 by a synthetic cohort of 100 people to whom the observed age-specific percentages never married applied.

Step 3: ***Add 1500 to the answer at Step 2***. This number is the number of person-years lived never married by the synthetic cohort ***before*** age 15 (100 × 15). We thus now have total person-years lived never married before age 50.
Step 4: ***Average the percentages never married at ages 45–49 and 50–54***. This yields an estimate of the percentage still single at exact age 50.
Step 5: ***Multiply the figure obtained at Step 4 by 50***. This gives person-years lived never married up to exact age 50 by members of our synthetic cohort still single at that age.
Step 6: ***Subtract the answer at Step 5 from that at Step 3***. This gives person-years lived never married up to exact age 50 by members of our synthetic cohort who married before that age.
Step 7: ***Subtract the answer at Step 4 from 100***. This gives the number out of our synthetic cohort who married by exact age 50.
Step 8: ***Divide the answer at Step 6 by that at Step 7***. This apportions person-years lived never married before age 50 by members of our synthetic cohort who married before age 50 among cohort members who married by exact age 50. It is, in other words, the ***average*** person-years lived never married by these individuals, or the ***singulate mean age at marriage***.

Note, however, that ***if you are dealing with a population in which some first marriages take place at ages younger than 15*** (i.e., the percentage of 10–14 year-olds never married is less than 100), two modifications must be made to this recipe. ***First***, at Step 1, add in the percentage never married at ages 10–14 as well (i.e., add percentages never married at ages 10–14 to 45–49, rather than at ages 15–19 to 45–49). ***Second***, at Step 3 add 1000, not 1500, to your answer at Step 2. There are now only 10 years of life, not 15, through which our synthetic cohort lives without being at risk of marrying. We take age group 10–14 into account at Step 1, where we deal with ages at which first marriage ***does*** occur, rather than at Step 3, where we deal with ages at which it does ***not*** occur.

As an example, consider Table 5.3, which shows the population of Bangladesh in 1974 by marital status, age and sex. We can use these data to obtain the age-specific percentages never married for females and males that are given following Table 5.3, which we will use to calculate singulate mean ages at marriage for women and men using the recipe outlined above.

Because in Bangladesh in 1974 some marriages of both females and males had taken place at ages 10–14 we must modify step 1 in accordance with the note given following the SMAM recipe, and add, for each sex, percentages never married for ages 10–14 to 45–49.

The singulate mean ages at marriage for Bangladeshi women and men in 1974 were 15.93 years and 23.96 years, suggesting an average age difference between husbands and wives of around eight years. Obviously the SMAM for females at this time was especially young, and conducive to high fertility. Subsequent calculations show the female SMAM rising to 17.90 by 1991 and to 19.34 by 2011, with the male SMAM rising to 24.85 and then 25.01 at these dates. So female ages at first marriage have risen more rapidly than have those for men and the age difference between

Table 5.3 Marital status by age and sex, Bangladesh 1974

Age group	Never married	Married	Widowed	Divorced	Total
Females					
Under 10	12,576,557	0	0	0	12,576,557
10–14	3,794,840	370,578	5,125	23,743	4,194,286
15–19	676,761	1,983,740	26,560	77,496	2,764,557
20–24	80,869	2,320,267	43,982	50,404	2,495,522
25–29	21,929	2,391,576	67,516	31,047	2,512,059
30–34	11,267	1,892,513	107,461	16,188	2,027,429
35–39	7,648	1,599,096	161,732	11,204	1,779,680
40–44	6,872	1,232,000	266,850	8,239	1,513,961
45–49	3,639	824,457	264,319	5,046	1,097,461
50–54	3,659	665,865	431,272	4,173	1,104,969
55–59	1,853	305,543	265,882	2,568	575,846
60–64	2,251	272,159	486,487	2,463	763,360
65 and over	4,916	208,707	785,640	2,077	1,001,340
Males					
Under 10	12,614,426	0	0	0	12,614,426
10–14	4,952,787	32,817	745	180	4,986,529
15–19	2,912,110	234,682	5,052	1,909	3,153,753
20–24	1,451,164	944,778	15,409	4,827	2,416,178
25–29	529,113	1,795,054	23,614	5,656	2,353,437
30–34	115,533	1,893,830	22,347	4,074	2,035,784
35–39	44,098	1,965,393	22,336	2,960	2,034,787
40–44	26,095	1,688,583	27,303	2,563	1,744,543
45–49	15,178	1,334,676	27,717	1,723	1,379,294
50–54	12,913	1,230,463	38,627	1,694	1,283,697
55–59	6,298	739,760	29,211	758	776,027
60–64	7,846	855,691	54,593	1,154	919,284
65 and over	12,345	1,211,675	147,891	1,076	1,372,987

Age group	Percentage of females never married	Percentage of males never married
Under 10	100.000	100.000
10–14	90.476	99.323
15–19	24.480	92.338
20–24	3.241	60.060
25–29	0.873	22.483
30–34	0.556	5.675
35–39	0.430	2.167
40–44	0.454	1.496
45–49	0.332	1.100
50–54	0.331	1.006
55–59	0.322	0.812
60–64	0.295	0.853
65 and over	0.491	0.899

(continued)

Table 5.3 (continued)

	Step number	Females	Males
1	Add percentages never married at ages 10–14 to 45–49	120.842	284.642
2	Multiply answer at step 1 by 5	$120.842 \times 5 = 604.210$	$284.642 \times 5 = 1{,}423.210$
3	Add 1,000 to answer at step 2 (in accordance with the note given following the SMAM recipe)	$604.210 + 1{,}000 = 1{,}604.210$	$1{,}423.210 + 1{,}000 = 2{,}423.210$
4	Average percentages never married at ages 45–49 and 50–54	$(0.332 + 0.331)/2 = 0.3315$	$(1.100 + 1.006)/2 = 1.0530$
5	Multiply step 4 answer by 50	$0.3315 \times 50 = 16.575$	$1.0530 \times 50 = 52.650$
6	Subtract answer at step 5 from answer at step 3	$1{,}604.210 - 16.575 = 1{,}587.635$	$2{,}423.210 - 52.650 = 2{,}370.560$
7	Subtract answer at step 4 from 100	$100 - 0.3315 = 99.6685$	$100 - 1.0530 = 98.9470$
8	Divide answer at step 6 by answer at step 7	$1{,}587.635/99.6685$	$2{,}370.560/98.9470$
		SMAM = 15.93	***SMAM = 23.96***

husbands and wives has narrowed. Teenage marriage for females is, however, still common in Bangladesh and significant scope still exists for further lowering fertility by increasing female ages at marriage.

It is sometimes claimed that the SMAM and e^o_0 in a gross nuptiality table are equivalent. This, however, is ***only true if marriage is universal by exact age 50***. The SMAM is a measure of exposure in the single state by persons ***who marry*** before they turn 50. By contrast e^o_0 also takes into account exposure in the single state, up to exact age 50, by those who ***fail*** to marry by that point in their lives (i.e., by (i) those who marry for the first time ***after*** they turn 50 and (ii) those who ***never*** marry). Thus, except where everyone marries before their 50th birthday, e^o_0 is biased upward in comparison to the SMAM and should be higher than it.

Consensual Partnering

The rise of consensual partnering in developed populations as an integral component of the 'second demographic transition' has greatly complicated the demographic analysis of 'partnering' in the context of understanding fertility change. Where formerly in most countries the commencement of co-residence and regular sexual relations, and hence susceptibility to conception, largely coincided with formal marriage in a climate devoid of other than relatively unreliable methods of contraception, nowadays consensual partnering and the capacity modern contraceptive methods (often backed by ready access to induced abortion) have provided to avoid parenthood for extended periods (and even completely) have substantially loosened the link between marriage and becoming parents. The picture is further complicated by the range of perceptions parties to consensual unions have of their unions, both cross-sectionally and over time. Along both dimensions they range in their levels of commitment from convenient arrangements for regular sexual gratification and/or securing domestic services and/or reducing living costs through modern approaches to 'going steady' and forms of engagement to marriage substitutes, with no guarantee that both parties to a union share the same perception, and hence degree of commitment, at a given point in time.

Consensual unions are generally not registered in the way marriages often are, so cannot be studied via registration data. Indeed while marriages can usually be dated because of the ceremony that surrounds them it can be difficult to date the commencement of consensual unions, which often evolve from couples initially staying the night with one another, then doing this with increasing regularity until the point of continuing to maintain two residences is questioned. Some censuses nowadays collect data on cohabitation that allows cohabitors and married persons to be compared. Ideally separate questions are asked on legal marital status and living arrangement (whether living with a spouse, living with a cohabiting partner, or not in a coresidential relationship), so that the extent to which cohabitors have been never and ever previously married can be examined, but with the pressure to limit the number of questions in censuses a single question that compromises such an investigation is often used.

More detailed studies of cohabitation have been based on survey data, and particularly on what are known as event history data. These provide detailed histories of dates/ages of entry to and exit from cohabiting unions and marriages (as well as of other events like births of children, migratory moves, changes of employment status, etc.). They are analysed using ***event history analysis*** techniques, a battery of regression models that overcome shortcomings (centred on ***censoring*** and the reality that ***explanatory variables can be time-varying***) of standard multiple regression for analysing event history data (Allison 1984; Yamaguchi 1991; Blossfeld and Gotz, 2001; Blossfeld et al. 2007; Brostrom 2011; Mills, 2011). It is not proposed to explore these techniques here, but they include such methodologies as ***logit regression***, ***proportional hazards modelling*** and ***accelerated failure time models***.

Analyses of partnering that use event history data focus on what is commonly termed ***union formation*** – that is, entry into either cohabiting unions or marriage. This can have the unfortunate effect of ***clouding the complexity of cohabitation as an institution***, tending to equate commencements of cohabiting unions with boyfriends or girlfriends with those of marriages as life cycle events of comparable moment and commitment, which in many instances they are not. Equally the dissolution of cohabiting unions and marriages may be of dissimilar moment – the ending of boyfriend-girlfriend relationships of perhaps limited duration that were never more than tentatively committed as opposed to separations that presage divorce. Demography still has a way to go to deal adequately with the inherent complexity of consensual partnering. It serves different purposes for different participants, and those purposes may differ for the parties to individual relationships and may change over time.

References

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Newbury Park/London/New Delhi: Sage.

Barrow, C. (1996). *Family in the Caribbean: Themes and perspectives*. Kingston: Ian Randle Publishers.

Barrow, C. (1998). Caribbean masculinity and family: Revisiting 'marginality' and 'reputation'. In C. Barrow (Ed.), *Caribbean portraits: Gender ideologies and identities* (pp. 339–358). Kingston: Ian Randle Publishers and Centre for Gender and Development Studies, University of the West Indies.

Blossfeld, H.-P., & Rohwer, G. (2001). *Techniques of event history modelling: New approaches to causal analysis*. Mahwah: Lawrence Erlbaum.

Blossfeld, H.-P., Golsch, K., & Gotz, R. (2007). *Event history analysis with stata*. Mahwah: Lawrence Erlbaum.

Brostrom, G. (2011). *Event history analysis with R*. Bosa Roca: Taylor & Francis.

Carmichael, G.A. (1988). *With this ring: First marriage patterns, trends and prospects in Australia.* Canberra: Department of Demography, Australian National University and Australian Institute of Family Studies.

International Union for the Scientific Study of Population. (1958). *Multilingual demographic dictionary: English section*. New York: United Nations Department of Economic and Social Affairs.

Mills, M. (2010). *Introducing survival and event history analysis*. London: Sage.

Rubenstein, H. (1980). Conjugal behaviour and parental role flexibility in an Afro-Caribbean village. *Canadian Review of Anthropology and Sociology, 17*(4), 330–337.

Shryock, Henry, S., Siegel, Jacob S. & Associates (1973) *The methods and materials of demography*. 2nd Printing (revised). Washington, DC: U.S. Bureau of the Census.

Van de Kaa, D. (1987). Europe's second demographic transition. *Population Bulletin, 42*(1), 1–57.

Wunsch, G. J., & Termote, M. G. (1978). *Introduction to demographic analysis principles and methods*. New York: Plenum Press.

Yamaguchi, K. (1991). *Event history analysis*. Thousand Oaks: Sage.

Chapter 6
Analysis of Fertility

The Importance of Fertility Analysis in Demography

The analysis of human fertility is of central importance in demographic analysis. In the first place, births are a vital component of the population balancing equation introduced in Chap. 1, and thus a vital element in the study of population change. More particularly they are, on a world scale, the ***single*** growth element in the balancing equation, and concern over population growth arguably has been the main stimulus to demography's development as a discipline. Efforts to slow population growth typically concentrate on attempting to reduce fertility. Deliberately increasing mortality is not an acceptable option, and while restricting immigration sometimes offers short-term potential, promoting emigration usually is neither a feasible nor a long-term solution either. The need to monitor efforts to reduce population growth by reducing fertility highlights the importance of being able to measure fertility, fertility change, and the sources of that change.

Fertility change also is intimately linked to the processes of social and economic development. The movement of women away from traditional roles which see them regarded more or less exclusively as bearers and nurturers of children, for example, generally is reflected in fertility decline and change in the age pattern of childbearing. The provision of health services, especially those directly linked to reproduction, and development-induced changes to household economies may have similar impacts. Debate over the causes and consequences of fertility change, and in particular fertility ***transition*** from high to low levels, has generated a substantial, hotly contested theoretical literature. Indices and approaches to analysis discussed in this chapter are among the tools that have facilitated this debate, in some instances having been developed specifically for it.

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_6

Some Terminology

It is important initially to become familiar with certain terms used in the fertility analysis literature. The first is the word ***fertility*** itself, which in everyday, including medical, usage refers to a person's ***biological capacity*** to bear or father children. In demography, however, a person's fertility is their ***actual reproductive performance***, and it is this that we are seeking to measure when we talk about 'measuring fertility'. Demographers' term for the biological capacity to bear children is ***fecundity***, and a woman capable of bearing a child is said to be ***fecund***. The biological ***in***ability to bear children is referred to as ***infecundity*** or ***sterility***, and a woman unable to bear a child is said to be ***infecund*** or ***sterile*** (the latter word, but not the former, also being used to describe men who are biologically unable to become fathers). In everyday usage 'fertility' is the opposite of 'sterility', but in demography only the latter word retains its everyday meaning. Arguably if referring to the 'fertility' ***of men*** there could be doubt as to whether one meant reproductive performance or biological capacity to reproduce (since one would not refer to the 'fecundity' of men), but while some would have it otherwise, fertility analysis in demography conventionally focuses exclusively on women, so this tends to be a non-issue. That said there has, of late, been interest in measuring male involvement in reproduction, and some of this literature has used the term ***'paternity'*** as the male equivalent of female 'fertility' (Carmichael 2013).

If all this seems confusing, note that the French word for 'fertility' as understood by demographers is 'fécondité', and that for 'fecundity' is 'fertilité'. Be wary, therefore, should you have occasion to read demographic literature written in French (as much important demographic literature is).

Used alone the word 'sterility' usually is assumed to refer to an irreversible condition, but a distinction is sometimes made between ***temporary*** sterility and ***permanent*** sterility. Among women a distinction may also be made between ***primary*** sterility (where a woman has ***never*** been able to have a child) and ***secondary*** sterility (where the biological inability to bear children develops after a woman has already had one or more children).

Reference was made in Chap. 1 to ***menarche*** and ***menopause*** as respectively the points in the life cycle at which a woman first acquires, and finally loses, the biological capacity to conceive and bear children. More specifically, menarche marks the onset, and menopause the final cessation, of ***menstruation***, a woman's experience of menstrual periods. Menarche is a discrete event, occurring with the occurrence of the first menstrual period, but 'menopause' often refers to an ill-defined ***period*** of the life cycle over which menstruation gradually ceases.

Another word commonly encountered in analyses of fertility is ***amenorrhoea***. It means a temporary absence of menstruation. This occurs, of course, during a pregnancy (***pregnancy*** amenorrhoea), but of greatest interest to demographers is the extent to which it occurs following a ***confinement*** (the process of actually giving birth), when it is known as ***post-partum*** amenorrhoea. The length of the period of post-partum amenorrhoea is highly variable, in particular as a function of whether or not, and for how long, a baby is breastfed. This variability and its relation to breastfeeding are of interest to demographers in populations practising little or no efficient contraception, because they affect how quickly after a birth a woman is at risk of becoming pregnant again, and therefore the interval between births. Promoting breastfeeding is one way of attempting to reduce fertility (as well as of improving infant and maternal health and survival chances) in such populations, because in increasing the period of post-partum amenorrhoea it lengthens the average interval between births, thereby reducing the number of children a woman bears over a given life cycle phase.

Fertility data distinguish between ***live births*** and ***stillbirths***. The World Health Organization defines a live birth as occurring when a foetus, whatever its gestational age, exits the mother's body and subsequently ***breathes or shows any sign of life***, such as voluntary movement, a heartbeat or pulsation of the umbilical cord, for however brief a time and regardless of whether the umbilical cord and placenta are intact (WHO 1993). A stillbirth occurs when a foetus has died in the uterus. The basic WHO definition of a foetal death is the intrauterine death of any conceptus at any time during pregnancy, but legal definitions often require the foetus to have attained a prescribed gestational age or birth weight to distinguish foetal deaths (stillbirths) from miscarriages. Variability in the minimum gestational age or birth weight prescribed in such definitions means you should always check the precise definition in use when dealing with data on foetal deaths/stillbirths.

Basic Fertility Measures

In keeping with the discussion of Chap. 3, ***fertility may be measured and analysed in either period (cross-sectional) or cohort terms***. In further keeping with that discussion, the former approach deals with births arranged vertically through the Lexis diagram (see Fig. 3.7), and the latter with births arranged diagonally through it and occurring normally to birth cohorts of women (although analysis in terms of other types of cohorts – e.g., marriage cohorts – is also conceivable). Period analyses of fertility have the advantage of being up to date, with a new year's measures added as each year's data are released, and are useful for forecasting purposes. Cohort analyses reflect the experience of actual birth cohorts, and are not distorted by transient period, or cross-sectional, effects. The more common and readily calculated fertility measures tend to be period measures, several of which were introduced in Chap. 1 in illustrating the more general discussion there of ***types*** of demographic measures.

Recapping, the ***crude birth rate*** (CBR) is the most basic ***relative*** measure of fertility. Defined as the ratio of live births during year y to the mean (or mid-year) total population during year y, multiplied by 1,000 (see Chap. 1, Eq. 1.7), it is not a 'true' rate, because its denominator is not even a moderately accurate measure of the population at risk of giving birth. It includes men, and also women who are both too young (yet to reach menarche) and too old (post-menopausal) to be able biologically to bear children. Thus the CBR is for use when no more refined data are available, typically for broad international comparisons. It is, however, a better summary measure of fertility than is the crude death rate of mortality, despite the latter being a true rate. The proportion of total population who are females of childbearing age tends to be less variable than does the proportion who are in the high mortality age groups. This means that the sort of situation encountered in comparing CDRs for Australia and Malaysia in Chaps. 1 and 2, where the ***underlying*** differential in mortality levels was so distorted by differences in age structure as to be ***reversed***, is less likely to occur.

A brief digression is in order while we are discussing the relative merits of the CBR and the CDR. You will recall from the discussion of the population balancing equation in Chap. 1 that population change over time is the sum of two components: natural increase and net migration. The difference between the CBR and the CDR (i.e., CBR – CDR) gives a measure known as the ***crude rate of natural increase***. Let us, though, return to our consideration of basic fertility measures.

In endeavouring to obtain measures of fertility more refined than the CBR the broad refinement principle outlined in Chap. 1 is applied. We seek to close in on the ***population at risk*** – by eliminating from the denominator those who could not possibly give birth, recognizing that certain personal characteristics make a woman more, or less, likely to give birth, or both. As previously discussed (see Chap. 1, Eq. 1.10), a first refinement of the CBR is to calculate the ***general fertility rate*** (GFR) by restricting the denominator to the female population aged 15–49 (i.e., to those persons who, broadly, are biologically capable of bearing children). This denominator is still only a rough approximation of the population at risk, since females may be fecund before age 15 and beyond age 49, and some of those aged 15–49 will be ***in***fecund (either involuntarily through having yet to reach menarche, already passed menopause or suffering from some dysfunction of the reproductive organs, or voluntarily through having been surgically sterilized). It could also be argued that some fecund women aged 15–49 are at no risk of giving birth because they are not sexually active or their sexual partners are sterile. Then, too, risk occurs in varying degrees among those actually exposed to it. Some women are more fecund or have sexual partners with more potent semen than others, and so conceive more readily; differing patterns of partner absence and frequencies of coitus affect risk, as do varying tendencies to miscarry, have stillbirths and have menstruation interrupted by breastfeeding or nutritional deficiencies; and increasingly, use of contraception has facilitated self-definition of risk. The concept ‘population at risk’ thus becomes very complicated when thought about in detail. But the denominator of the GFR (sometimes restricted to females aged 15–44 instead of 15–49) clearly represents a much more reasonable population at risk of giving birth than does the total population.

A second refinement discussed in Chap. 1 was to recognize that married women are more likely to be sexually active and to be consciously trying to have children than are unmarried women, so that there was a case for splitting the GFR by marital status into a (general) ***marital fertility rate*** (see Chap. 1, Eq. 1.11) and a (general) ***non-marital fertility rate*** (sometimes also called the ***illegitimacy rate***, although this terminology is generally frowned upon these days) (see Chap. 1, Eq. 1.12). A more fundamental refinement, however, is by ***age***, since within the reproductive age range (15–49 years) a variety of both biological and social factors make childbearing more likely at some ages than at others. Comparisons of GFRs thus are susceptible to distortion due to differences in age structure within the reproductive range. The calculation of ***age-specific fertility rates*** (or ***birth rates***) (ASFRs) is one way of focusing on this age variability. Techniques of standardization and decomposition can be applied to the GFR as to other demographic indices to control for differences in age structure, but in fertility analysis the detail offered by age-specific rates often is inherently of interest and is not too difficult to digest given that the female reproductive ages span only seven five-year age groups. Age-specific rates are also the building blocks of the most widely used fertility measure – the ***total fertility rate (or ratio)*** (TFR – see below).

If data on births by age of father are available, it is possible to calculate ASFRs (or age-specific ***paternity*** rates – ASPRs) for males. Usually, however, they are computed only for females. The equation for a (female) ASFR is:

$$\mathrm{ASFR_x} = (\mathrm{B_x}/\mathrm{P_x})\,1{,}000 \tag{6.1}$$

Where x denotes an age group; $\mathrm{B_x}$ = live births to women aged x during year y; $\mathrm{P_x}$ = the mean (or mid-year) female population aged x in year y.

While in theory the age range 15–49 years can be divided up in many ways for the purpose of calculating ASFRs, in practice rates most commonly are calculated for five-year age groups, sometimes for single-year age groups. ASFRs for five-year age groups 15–19 to 45–49 years often are used to contrast fertility patterns of discrete populations (Does fertility peak at different ages? How do relativities in fertility levels at different ages compare?), and to compare patterns for individual populations over time (Was fertility increase or decline concentrated at particular ages?).

Both five-year and single-year ASFRs also are used to obtain ***the TFR***. Relevant equations were given as Eqs. 1.19 and 1.20, respectively, in Chap. 1. As indicated at that point, the TFR measures the number of children the average woman would have during her life ***assuming*** she experienced at each reproductive age the ASFR for that age recorded during the year for which the index is calculated. It is a ***very widely used*** measure of fertility. Demographers often summarize fertility change, especially fertility decline, by quoting before/after TFRs (e.g., the TFR fell from 6.4 to 3.7), and frequently judge the fertility level in a population with reference to a TFR of from about 2.1 (for low mortality populations) to about 2.4 (for high mortality populations). These are approximate levels of fertility required for ***long-term replacement*** – i.e., levels that, maintained indefinitely along with existing levels of mortality, would see a population exactly replace itself, the new generation being exactly the same size as the older, parental generation. The figures of 2.1–2.4 comprise one child to replace the woman, one to replace her male partner, and 0.1–0.4 as a contribution to the additional fertility required to replace those unable to participate, or participate fully, in the replacement process because of premature death. Clearly the lower the mortality in a population the lower this ‘per woman’ contribution needs to be. In Australia, which has low mortality, the precise value of the replacement level TFR currently is about 2.06, but 2.1 is a commonly used general guideline for developed country populations. A TFR below long-term replacement, if maintained, signals eventual population decline, although the onset of decline invariably is delayed for quite some time by ***momentum*** for increase built into the population’s age structure by past, above replacement, fertility; a TFR above long-term replacement level implies a growing population.

While specific fertility rates are most commonly computed specific for age, there are other possibilities. Marital status already has been mentioned. Fertility levels, differentials and trends for the married and unmarried components of populations often are very different, and changing or different marital status compositions can be important explanations of fertility trends or differentials as indexed by measures like the CBR or GFR. In addition to the (general) marital and non-marital fertility rates already discussed, ***age-specific marital fertility rates*** (ASMFR) and ***age-specific non-marital fertility rates*** (ASNMFR) can be valuable in examining such issues. Their equations are:

$$\mathrm{ASMFR_x} = \left(\mathrm{B_{(m,x)}}/\mathrm{P_{(f,m,x)}}\right) 1{,}000 \tag{6.2}$$

Where x denotes an age group; $B_{(m,x)}$ = marital live births to women aged x during year y; $P_{(f,m,x)}$ = the mean (or mid-year) married female population aged x in year y.

And:

$$\mathrm{ASNMFR_x} = \left(\mathrm{B_{(n,x)}}/\mathrm{P_{(f,u,x)}}\right) 1{,}000 \tag{6.3}$$

Where x denotes an age group; $B_{(n,x)}$ = non-marital live births to women aged x during year y; $P_{(f,u,x)}$ = the mean (or mid-year) unmarried female population aged x in year y.

Of course the contemporary trend to cohabitation rather than marriage for significant chunks of the life course in developed populations raises the issue of exposure to the risk of bearing children while cohabiting. Cohabitors of childbearing age are obviously likely to be regularly sexually active as married couples are, and so exposed to the risk of pregnancy. They nonetheless tend to be treated as unmarried, and their fertility is captured by the various non-marital fertility rates. In theory, if mid-year estimates of population cohabiting and annual counts of non-marital births to cohabiting couples were available it would be possible to compute separate specific fertility rates for cohabiting (as opposed to non-coresident) unmarried women, but thus far requisite data are rarely available.

Order-specific fertility rates (OSFR) and ***age-order-specific fertility rates*** (AOSFR) are another two types of specific fertility rates which pay attention to ***birth order*** (whether a birth is the mother’s first, second, third, etc. live birth).

Birth order and its companion attribute, a woman's ***parity***, are important variables in fertility analysis, since as the number of children a woman has increases so do the likelihoods (a) that she will wish to take steps to avoid further births and (b) of her physiological capacity to have further children being impaired by previous childbearing and/or advancing age. The relevant equations are:

$$\mathrm{OSFR_i} = \left(\mathrm{B_i}/\mathrm{P_{(f,15-49)}}\right) 1{,}000 \tag{6.4}$$

Where i denotes a birth order; $\mathrm{B_i}$ = live births of order i during year y; $\mathrm{P_{(f,15-49)}}$ = the mean (or mid-year) female population aged 15–49 in year y.

$$\mathrm{AOSFR_{(x,i)}} = \left(\mathrm{B_{(i,x)}}/\mathrm{P_{(f,x)}}\right) 1{,}000 \tag{6.5}$$

Where x denotes an age group; i denotes a birth order; $\mathrm{B_{(i,x)}}$ = live births of order i to women aged x during year y; $\mathrm{P_{(f,x)}}$ = the mean (or mid-year) female population aged x in year y.

Note that in these two equations the denominators are respectively ***all*** women of reproductive age and ***all*** women aged x; they are ***not*** women ***of parity i-1*** and the relevant age who, ignoring the complication of multiple births, are the women strictly at risk of producing births of order i. It follows that the sum of OSFR over all birth orders is the GFR, while the sum of AOSFR over all birth orders for age group x is the ASFR for age group x. The value of these indices is thus that they allow the GFR and ASFRs to be partitioned by birth order. Such exercises may aid understanding of the mechanics of fertility change over time – for example, the extent to which overall change reflects change in the frequency of higher order births.

You do need to be wary of registration data on births by birth order (or mother's parity). They can be seriously inaccurate. In Australian birth registration data, for example, births by mother's parity have historically been based on questions on birth registration forms (the precise wording of which varies across the States and Territories that have jurisdiction for vital registration) asking for details of previous children ***of this marriage*** or, more recently, ***of this relationship*** (Carmichael 1986; Corr and Kippen 2006). So they exclude children of ***previous*** marriages/relationships, and when 'this marriage' was the referent, previous issue of the mothers of non-marital births, and hence the parities of those births, were ignored altogether. Parities have thus been systematically understated through such things as a previous non-marital birth by another man being ignored, or a woman divorcing after having children, remarrying, and then returning to parity one when bearing a first child by her second husband. The Australian Bureau of Statistics (ABS) has been trying for some years to persuade State/Territory registrars of births to collect parity data that reflect a woman's ***complete*** fertility history, but the alleged sensitivity of requiring disclosure of the details of previous births a current partner may be unaware of has prevented this happening in some States. As of 2007 the

ABS began publishing parity data of this type for States/Territories that collected it, but this excluded two of the largest States, Queensland and Victoria, for which such data continue to be impaired. Current national data are thus hybrid.

Fertility rates specific for age, marital status and birth order are but three examples of many specific rates that might be useful in fertility analysis. Rates specific for variables such as educational attainment, ethnic group, occupation, contraceptive use, income, etc. are also conceivable. Their equations take the form 'live births to women with the relevant attribute divided by the mean, or mid-year, female population with that attribute, multiplied by 1,000', and invariably they could usefully be calculated specific also for age.

Fertility Measures from Censuses and Surveys

Measures discussed in the previous section require two types of data. They require data on ***population***, classified by age, sex and possibly other variables, and data on ***live births***, classified by age of mother and possibly other variables. The former typically derive from a census or from census data adjusted to the date of interest for intervening vital events; the latter normally derive from vital registration. Many countries, especially developing countries, do not, however, have reliable vital registration data, and in this circumstance fertility measures which require only census or survey data are widely used. The focus in this section is on measures which can be obtained directly from such sources, as opposed to those obtained from them using what are known as the ***methods of indirect estimation*** (or ***indirect methods***) (United Nations 1983).

Two types of fertility data commonly are collected by censuses: Data on ***children ever born***, and on ***births within the previous 12 months*** (so-called 'current fertility'). These types of data may also be gathered through surveys, although surveys may collect other types of fertility data as well, the most detailed of all being full pregnancy histories.

Measures Based on Children Ever Born

Data on children ever born (CEB) can be used to construct estimates of standard fertility measures (like the TFR) using indirect methods. However, these methods require ***either*** data from two successive censuses ***or*** acceptance of an assumption that fertility has not changed in the recent past. Assumptions of lack of change in demographic parameters often are made in demographic analysis for populations for which data are poor, and many techniques of analysis are only modestly sensitive to this type of assumption and therefore quite robust. However, methods of ***fertility*** analysis which assume lack of change ***usually are VERY sensitive to that assumption, and should be used with extreme caution***. Methods of indirect estimation do not, though, fall within the scope of the present discussion. Data on CEB are mentioned because they also yield other useful summary measures.

One such measure is the ***proportion of women childless at each age***. This, at younger ages, provides information about the speed with which childbearing begins in a population, but a more common use is as a measure of the extent of infecundity at ages above exact age 40. When used for this purpose in relation to populations in which marriage is not universal, calculations obviously should be restricted to ever married women.

A more widely used measure using CEB data is the ***age-specific mean number of children ever born***. For any age group this mean summarizes the cohort's fertility history, and its main use is in ***differential fertility analysis*** – i.e., in comparisons of the fertility experience of subgroups of women within the population under study. Provided they are reliable, mean CEB data are likely to facilitate a much wider range of fertility comparisons than are 'standard' measures like age-specific and total fertility rates. The detailed cross-tabulations of births needed to calculate the latter for many subgroups frequently are not available in registration data, whereas mean CEB data, deriving from a single data source and one apt to include a wide range of personal attributes besides CEB, generally offer a much wider range of comparison options (e.g., by education, labour force status, rural-urban residence, ethnicity, religion).

The qualification 'provided they are reliable' made in respect of mean CEB data must not, however, be lightly brushed aside. Regrettably, CEB data frequently have deficiencies. At younger ages it is not uncommon to encounter large numbers of 'not stated' responses because of failure to obtain information for young unmarried women. Census enumerators may have felt uncomfortable asking the question of such women or have met frequent refusals to answer it, or in self-enumerations householders may have deemed the question irrelevant or inappropriate in respect of their unmarried daughters. A mean CEB calculation which ignores 'not stated' cases in this circumstance generally has a serious upward bias, since the great majority of those cases are likely to be childless. There are analytical techniques which attempt to compensate for this bias, but it is safest to work on the basis that the calculation and comparison of mean numbers of CEB should only be undertaken where the 'not stated' category is small. Often, in fact, CEB questions are not even asked of the never married, in which case mean CEB data obviously cannot be interpreted as if they applied to all women.

It is also sometimes found that mean CEB for older age groups (beyond, say, exact age 35 or 40) are less reliable than those for younger age groups. Possible reasons for this include the exclusion of deceased children and, in societies in which adoption is common, confusion between biological and social parenthood. Again, caution is called for. Age-specific mean numbers of CEB are nonetheless an extremely valuable type of demographic measure, especially for analyses of differential fertility focused on age group 30–34 (and perhaps the two adjacent age groups), at which data defects just canvassed tend to be least in evidence.

'Current Fertility' Measures

Asking about live births in the preceding 12 months is most common in censuses and surveys undertaken in less developed countries, because it yields data almost equivalent to birth registration data (and these are, of course, the types of countries least likely to have such data). In theory 'current fertility' data are in one sense superior to registration data; they facilitate calculation of standard fertility measures for the year preceding the census using a single data source, thereby eliminating problems of differential levels of enumeration in registration and census data. However, they have other problems of their own.

There are often problems of coverage. For example, respondents may not observe a precise 12-month reference period, births ending as infant deaths may be ignored, and distinguishing stillbirths from early neonatal deaths (and hence whether a live birth occurred) may be more problematic than ever. Also, births are classified by the mother's age at enumeration, which may be different from her age at the date of the birth. On average births occurred 6 months prior to the census or survey, so that when classified by mother's age in five-year groups, these age groups effectively are displaced 6 months. This need not deter direct calculation of a TFR, but this is not often done. Instead, current fertility data more typically are used as input for indirect methods – not always with certainty that the outcome will be an improvement on a direct calculation.

The Child-Woman Ratio

The measures discussed so far in this section often are regarded as part of the body of methods of indirect estimation, although there would seem to be nothing indirect about calculating the mean number of CEB for a female age cohort. But another type of fertility measure that undeniably is calculated directly from census data (though how direct a measure of fertility it is is another question) is the ***child-woman ratio***.

The child-woman ratio is different in not deriving from any type of fertility data as such. It requires only a census age-sex distribution, and is the ratio of children aged 0–4 to females aged 15–49, multiplied by 1,000 (see Chap. 1, Eq. 1.18). It is a somewhat rough measure, making no allowance, in a comparative context, for differences in female age structure across the reproductive ages, for differently shaped age-specific fertility distributions or for different levels of child survival. It is perhaps best regarded as giving a reasonable approximate guide to differences in fertility levels in situations where its limited data requirements commend it.

Measures of Reproduction

The concept of ***reproduction***, or ***reproductivity***, as conventionally used in demography, has to do with the extent to which a cohort of women (either real or synthetic) – a ***generation*** of women – replaces, or reproduces, itself (by having ***daughters***). Its analysis led Alfred Lotka in the early twentieth century (Leonard Euler independently pre-empted Lotka as early as 1760) to the concept of a ***stable population***, which underpins several techniques of demographic analysis today. Euler and Lotka demonstrated that a closed population subjected indefinitely to fixed schedules of age-specific fertility and mortality rates ultimately assumed a fixed, or stable, age distribution (i.e., an age distribution with a constant ***shape***) and a constant rate of natural increase. Stable population theory will be addressed in Chap. 8. Reproduction rates can in theory be calculated for males or for both sexes combined as well as for females, but because the female reproductive span is shorter and better defined, and because mothers' ages are more likely to be known than fathers' in the case of non-marital births, rates linking daughters to mothers are the norm.

The Gross and Net Reproduction Rates

In distinguishing in Chap. 3 between the tempo and intensity of (real) cohort demographic processes, equations for two generalized measures of intensity were presented (as Eqs. 3.1 and 3.2), and it was noted that for the process of women giving birth to female children these equations yielded measures known as the ***net reproduction rate*** (NRR) and ***gross reproduction rate*** (GRR). These are the core measures of reproduction, and it is appropriate to give here explicit equations for them (in the real cohort context).

$$\mathrm{GRR} = \Sigma_{x=l,\omega}\left(b_{(f,x)}/{}_1F_x\right) \tag{6.6}$$

Where $b_{(f,x)}$ = the number of female births occurring at age x to a female birth cohort; ${}_1F_x$ = the number of survivors in the female birth cohort at age x (i.e., at the point in time when all cohort members are aged between exact ages x and x + 1); l = the youngest age at which any cohort member gave birth to a female child; ω = the oldest age at which any cohort member gave birth to a female child.

$$\mathrm{NRR} = \Sigma_{x=l,\omega}\left(b_{(f,x)}/N_0\right) \tag{6.7}$$

Where $b_{(f,x)}$ = the number of female births occurring at age x to a female birth cohort; N_0 = the original size of the female birth cohort at exact age 0; l = the youngest age at which any cohort member gave birth to a female child; ω = the oldest age at which any cohort member gave birth to a female child.

If we consider these equations applied in respect of a closed population it is easy to appreciate that the NRR measures the ***actual*** extent to which a cohort of women replaces itself with daughters, while the GRR measures the extent to which the cohort ***would*** replace itself ***if*** all members survived to the end of the reproductive period.

Equations 6.6 and 6.7 and the real cohort, closed population context in which they apply are helpful in ***understanding*** the GRR and NRR. There are, however, period, or synthetic cohort equivalents which use cross-sectional data, and these tend to be the GRRs and NRRs demographers ordinarily calculate. Thus, with single-year-of-age data the usual calculating equations for the GRR and NRR are:

$$\mathrm{GRR} = \Sigma_{x=1,\omega}\mathrm{ASFFR}_x = \Sigma_{x=1,\omega}\left(b_{(f,x)}/F_x\right) \tag{6.8}$$

Where ASFFR_x stands for the age-specific fertility rate at age x based on female births only; $b_{(f,x)}$ = the number of female births occurring to women aged x in year y; F_x = the mean (or mid-year) female population aged x in year y; 1 = the youngest age at which any woman gave birth to a female child in year y; ω = the oldest age at which any woman gave birth to a female child in year y.

And:

$$\mathrm{NRR} = \Sigma_{x=1,\omega}\mathrm{ASFFR}_x \,.\, L_x/l_0 = \Sigma_{x=1,\omega}\left(b_{(f,x)}/F_x\right)\ \ (L_x/l_0) \tag{6.9}$$

Where ASFFR_x stands for the age-specific fertility rate at age x based on female births only; L_x / l_0 is the life table survival ratio from birth to age x from a single-year-of-age life table for the female population in question; $b_{(f,x)}$ = the number of female births occurring to women aged x in year y; F_x = the mean (or mid-year) female population aged x in year y; 1 = the youngest age at which any woman gave birth to a female child in year y; ω = the oldest age at which any woman gave birth to a female child in year y.

Two difficulties may arise in applying Eqs. 6.8 and 6.9. First, data on births may not be refined by sex. Second, single-year-of-age data and/or a single-year-of-age life table may not be available. If data on female births are not available, approximations of values of ASFFR_x can be obtained by multiplying values of ASFR_x (age-specific fertility rates based on ***total*** births) by 100/205. This procedure makes use of the empirical regularity that a sex ratio of approximately 105 male births per 100 female births is observed in most populations, and if a more precise sex ratio at birth is available the denominator 205 can be adjusted accordingly.

Where single-year-of-age data or, in the case of the NRR, a single-year-of-age life table are not available there are modifications of Eqs. 6.8 and 6.9 which allow calculations to be based on data for five-year age groups and an abridged life table. Indeed, if undertaking calculations manually, these equations have considerable appeal even if their use is not essential, as they greatly reduce the quantity of mechanical computation. One minor drawback is that calculations normally are limited to the age range 15–49 years. Small numbers of female children born outside that range sometimes are assigned to the adjacent five-year age group (15–19 or 45–49), or if numbers are non-negligible it is possible to extend the summations in Eqs. 6.10 and 6.11 over age groups 10–14 and/or 50–54. Extension to include age group 10–14 is in particular not uncommon. The relevant equations based on five-year age groups are:

$$\mathrm{GRR} = 5\Sigma_{x=15-19,45-49}\left(b_{(f,x)}/F_x\right) \tag{6.10}$$

Where $b_{(f,x)}$ = the number of female births occurring to women aged x in year y; F_x = the mean (or mid-year) female population aged x in year y.

And:

$$\mathrm{NRR} = 5\Sigma_{x=15-19,45-49}\left(b_{(f,x)}/F_x\right)\ ({}_5L_i/5l_0) = \Sigma_{x=15-19,45-49}\left(b_{(f,x)}/F_x\right)\ ({}_5L_i/l_0) \tag{6.11}$$

Where i = the lower limit of age group x; ${}_5L_i/5l_0$ is the life table survival ratio from birth to the five-year age group whose lower limit is exact age i from an abridged life table for the female population in question; $b_{(f,x)}$ = the number of female births occurring to women aged x in year y; F_x = the mean (or mid-year) female population aged x in year y.

With Eqs. 6.10 and 6.11, as with Eqs. 6.8 and 6.9, if data on female births are not available the ASFFR_x element $b_{(f,x)}/F_x$ can be replaced by values of ASFR_x multiplied by 100/205 (or a more precise alternative sex ratio).

By now something about the GRR may have struck you. It is the TFR based on female rather than total births. It measures the number of ***daughters*** (as opposed to children) a woman would have through her life if experiencing the age-specific rates of having ***female*** (as opposed to total) births prevailing in the year for which it is calculated. ***This link between the GRR and the TFR allows a good approximation of either to be obtained from the other***, simply by using the appropriate multiplier based on the sex ratio at birth. Most commonly the TFR is used to estimate the GRR, the former being multiplied by 100/205 (or thereabouts). To convert in the opposite direction one would use the inverse of this multiplier (205/100).

Brief reference was made earlier to the concept of long-term replacement. Strictly speaking, ***replacement level fertility occurs when age-specific fertility rates based on female births and age-specific female mortality rates combine to produce a net reproduction rate of 1.0***. Thus achievement of replacement level fertility is slightly dependent on mortality conditions, and as previously suggested tends to coincide with a TFR of from about 2.1 in low mortality populations to about 2.4 in high mortality populations. In the latter, the survival of proportionately fewer women to the end of the reproductive period means that the average woman who ***does*** survive needs to make a larger contribution beyond the two children to replace herself and her male partner to compensate for the lost fertility of women who die prematurely. A NRR of 1.0 corresponds also with a GRR slightly ***higher*** than 1.0, and hence replacement level fertility can coincide with a GRR of from about 1.01 to about 1.20. The GRR is always higher than the NRR because it assumes ***all*** women survive through the reproductive period to experience the ASFFRs recorded during the year in question. The difference between the two measures tends to be slight when female mortality before age 50 is low, and larger where female mortality at those ages is higher. The ratio of the NRR to the GRR yields the ***reproduction-survival ratio***; the proportion of ***potential*** reproductivity that survives the effect of mortality. With a NRR of 1.0 implying ***exact replacement***; anything lower indicates ***sub-replacement*** fertility, and a higher value ***above replacement*** fertility.

As a measure of how large the generation produced by its predecessor will be in proportionate terms, the NRR has a readily understandable meaning. However, just as a below replacement TFR does not signal an immediate decline in population numbers, neither does a NRR below 1.0 give this indication. Normally, because of momentum for growth built into a population's age structure, NRRs below 1.0 would need to persist for some time before decline began. The NRR was a popular measure in the 1920s and 1930s, when low values for industrialized countries were interpreted as indicating levels to which fertility would ultimately fall, but it was labelled misleading when values suddenly rose again with the baby boom that occurred following the Second World War. In both instances the measure was being interpreted naively in real, rather than synthetic, cohort terms, but it has never recovered its earlier popularity (for a fuller discussion see Shryock, Seigel and Associates (1973: 534–535)). Requiring fewer data, ***the TFR has become by far the preferred and more widely used measure.***

Another important point to grasp about NRRs is that differences between them do not necessarily measure differences in rates of population growth, even if mortality conditions in populations being compared are identical. The reason is that ***net reproduction can take different lengths of time***, depending on the mean age of childbearing. A population that has high age-specific fertility at younger ages reproduces itself (creates the next generation) more quickly than one with the same TFR and NRR, but with age-specific fertility distributed more to older reproductive ages. The latter population is said to have a longer ***length of generation***.

Mean Length of Generation

The ***mean length of generation*** for a population is defined as the mean age of mothers at the birth of their daughters. It is an index designed to measure the speed with which a group of mothers replaces itself with another group of potential mothers. With single-year-of-age data and a relevant single-year-of-age life table the relevant equation is:

$$\mathrm{MLG} = \left[\Sigma_{x=1,\omega}\left(b_{(f,x)}/F_x\right)\ \ (L_x/l_0)\ \ (x+0.5)\right]/\left[\Sigma_{x=1,\omega}\left(b_{(f,x)}/F_x\right)\ \ (L_x/l_0)\right] \tag{6.12}$$

Where all elements mean the same as in Eq. 6.9.

With data for five-year age groups and a relevant abridged life table the relevant equation is:

$$\mathrm{MLG}= \left[\Sigma_{x=15-19,45-49}\left(b_{(f,x)}/F_x\right)\ ({}_5L_i/l_0)\ (m_x)\right]/\left[\Sigma_{x=15-19,45-49}\left(b_{(f,x)}/F_x\right)\ ({}_5L_i/l_0)\right] \tag{6.13}$$

Where m_x is the midpoint of age group x; other elements mean the same as in Eq. 6.11.

As with Eqs. 6.8, 6.9, 6.10, and 6.11, if data on female births are not available the $ASFFR_x$ element $b_{(f,x)}/F_x$ in Eqs. 6.12 and 6.13 can be replaced by values of $ASFR_x$ multiplied by 100/205 (or a more precise alternative sex ratio at birth). And in Eq. 6.13, summations can be extended over age groups 10–14 and/or 50–54 if non-negligible childbearing occurs at either or both of those ages.

Note that in Eqs. 6.12 and 6.13 the denominators are the NRR as given by Eqs. 6.9 and 6.11 respectively. MLG is a weighted mean of the ages at which births of daughters occur, ages being represented by the mid-points of the single-year or five-year age groups within which births occur, and the weights being, in effect, numbers of female births at those ages in the life table stationary population.

Values of MLG tend to fall in the range 25–30 years, and results radically outside this range should be carefully checked, although values over 30 for developed populations widely delaying childbearing into their 30s are not uncommon. A calculation of the MLG for Australia in 2011 using five-year age group data (which incorporates calculations of the GRR and NRR) is presented in Table 6.1. This table allocates known total births by age to female births by age using an observed overall ratio of female births to total births of 146,621/301,617 = 0.486116.

Thus Australia in 2011 had sub-replacement fertility. With the NRR at 0.922, replacement of females of reproductive age by daughters was 92.2 %. The mean length of generation was 30.5 years, having risen from 28.4 years in 1991. Thus the deferment of fertility to older ages over this 20-year period as many women prioritized other things (education, careers, travel, their social lives) through their teens and twenties had increased the MLG by 2.1 years.

Describing the Age Pattern of Fertility

In comparing the fertility of different populations or studying fertility change, demographers are interested in differences/changes in the ***tempo*** of fertility as well as in differences/changes in its level, or intensity. One way in which something of a focus on differences/changes in tempo is facilitated is through an examination of

Table 6.1 Calculation of gross reproduction rate, net reproduction rate and mean length of generation for Australia, 2011

Age group	m_x	$b_{(f,x)}$	F_x	$b_{(f,x)}\,/\,F_x$	${}_5L_i\,/\,l_0$	(5) × (6)	(2) × (7)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
15–19	17.5	5,515	706,860	0.007802	4.97293	0.03880	0.67900
20–24	22.5	20,305	788,193	0.025761	4.96647	0.12794	2.87865
25–29	27.5	40,912	817,086	0.050071	4.95958	0.24833	6.82908
30–34	32.5	46,516	766,950	0.060651	4.95101	0.30028	9.75910
35–39	37.5	27,042	791,706	0.034157	4.93878	0.16869	6.32588
40–44	42.5	5,903	800,496	0.007374	4.92008	0.03628	1.54190
45–49	47.5	296	777,690	0.000381	4.89109	0.00186	0.08835
				$\sum = 0.186197$		NRR = 0.92218	$\sum = 28.10196$
$\text{GRR} = 5\sum(5) = 5 \times 0.186197 = 0.930985$							
$\text{MLG} = \sum(8)\,/\,\sum(7) = 28.10196\,/\,0.92218 = 30.5$ years							

age patterns of fertility. The age pattern of a population's fertility is its distribution by maternal age, after controlling for differences by age in numbers of potential mothers.

Graphs of Age-Specific Fertility Rates

Age patterns of fertility are commonly examined by graphing age-specific fertility rates, usually for five-year age groups. When compared such graphs can very effectively highlight differences in the ages at which fertility peaks, the prominence of such peaks, and ages at which fertility differences or change are concentrated. They also provide a strong visual impression of differences/changes in the intensity of fertility.

Figure 6.1 plots age-specific fertility rates for several countries in the early 1990s chosen to illustrate the diversity of age patterns of fertility one can encounter. Bangladesh (TFR 4.45 children per woman) exhibits clearly the highest ASFRs at all ages, with peak fertility straddling the 20–24 and 25–29 age groups. Childbearing begins early, and fertility limitation receives scant attention while women are still in their twenties. Next, the former Soviet republic of Kazakhstan (TFR 2.66) is an example of a population where fertility is heavily concentrated at younger ages, with a distinct peak at ages 20–24 and fertility limitation practiced increasingly thereafter. The former German Democratic Republic (GDR), or East Germany, and the Federal Republic of Germany (FRG), or West Germany, then provide an interesting contrast. With TFRs of 1.59 and 1.41, respectively, their fertility levels are reasonably similar, but their age patterns of fertility are very different. The GDR pattern is similar to that of Kazakhstan, but at a lower level. Fertility is much higher at ages 15–19 and 20–24 than in the FRG, and peaks in the second of these age groups. By contrast FRG fertility occurs decidedly later, peaking at ages 25–29 at a

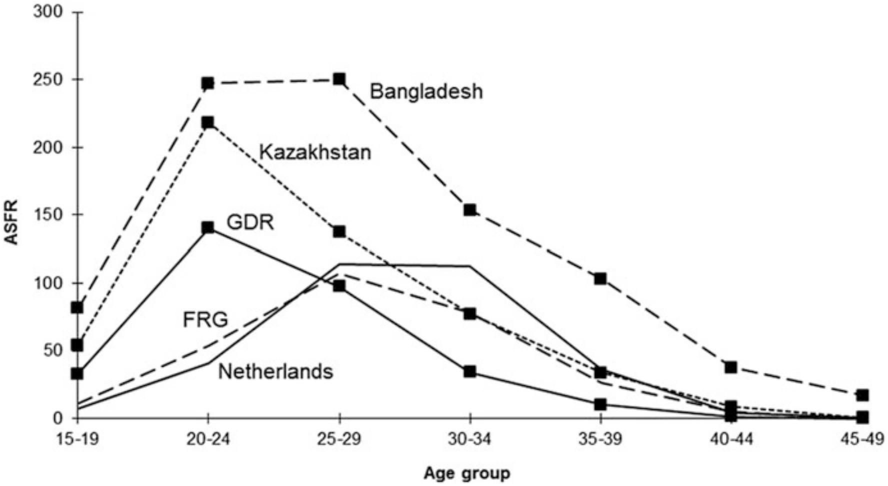


Fig. 6.1 Age-specific fertility rates for selected countries in the early 1990s (Source: United Nations demographic yearbook, 1994)

level only marginally above that observed in the GDR, but then being distinctly higher than in that country at ages 30–34 and 35–39. These differences reflect different options open to young adults in the two countries in the early 1990s. The FRG pattern is typical of more developed countries in recent times as the young have focused through their twenties on acquiring education, establishing careers, travelling, experimenting with relationships and having a good time. Australia's pattern in the early 1990s was similar, although its ASFRs were a little higher than those for the FRG. In the GDR prior to re-unification early marriage and childbearing were a more attractive path to independence given more restricted alternatives, but parlous economic circumstances then recommended a high level of fertility control beyond age 30. The Netherlands (TFR 1.58), finally, exhibits a very delayed age pattern of childbearing. Its flat peak across ages 25–29 ***and*** 30–34 was quite unusual in the early 1990s, although Australia would have a similar pattern 10 years later, and by 2011 age group 30–34 was its peak fertility age group. This can be seen in Fig. 6.2, which using Australia as an example, demonstrates change over time (since 1921) in the age pattern of fertility. TFRs represented by the fertility patterns shown are indicated in parentheses in the graph legends.

In 1921, with a TFR of 3.12 children per woman, fertility in Australia was highest at ages 25–29, but also very much higher at older ages than it would ever subsequently be, reflecting relatively late marriage and limited control over fertility within marriage. The 1934 pattern then captures the impact on fertility of the Great Depression, with fertility still highest at ages 25–29, but the TFR reduced by one-third and all age-specific fertility rates much lower. Clearly couples managed to control fertility to a substantial extent when economic circumstances demanded it, partly through deferring marriages (see Fig. 5.1 in Chap. 5). By 1945 fertility at

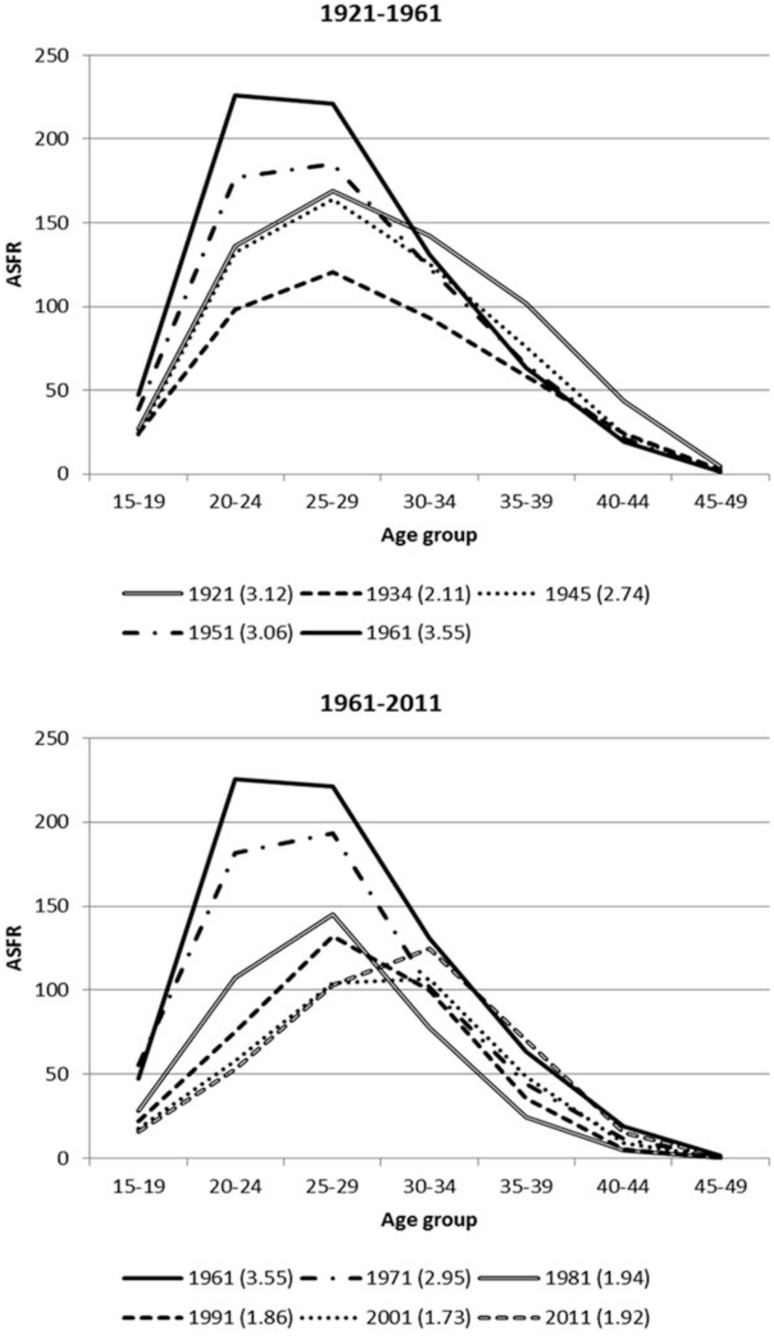


Fig. 6.2 Age-specific fertility rates: Australia 1921–2011 (Source: Australian Bureau of Statistics)

ages 15–19 to 25–29 had recovered to 1921 levels, but while that at older ages had also rebounded, it had not done so to earlier levels. Six years later the post-war baby boom was well under way, fertility having risen appreciably at ages 20–24 and substantially also at ages 15–19 and 25–29, and the TFR once again back above three children per woman. These trends continued through the 1950s until 1961, when the peak post-war TFR of 3.55 children per woman was recorded and age group 20–24 had taken over as the peak fertility age group. Couples were marrying earlier than ever before, not infrequently with the first child having already been conceived as a sexual revolution unfolded, and for most young women their twenties were heavily focused on having and raising children.

The year 1961 was the year oral contraception (the pill) first became available in Australia. By 1971 fertility had declined again (TFR 2.95) at all ages except 15–19 as the pill facilitated deferment of childbearing following marriage and avoidance of unwanted higher parity births by women in their thirties and forties, but was largely denied to young unmarried women other than in the shadow of planned weddings. By 1981 fertility decline had continued apace, with 15–19 year-olds by now incorporated into it, the TFR below replacement level at 1.94 children per woman, and with sterilization procedures now also firmly part of the arsenal for preventing unwanted higher parity births, fertility at ages 30–34 and 35–39 at lowest ever levels. The incorporation of teenagers into the trend reflected rising consensual partnering and use of modern contraception, and the opening up following a landmark 1971 legal ruling of access to induced abortion as the ultimate protection against unwanted childbearing. Beyond 1981 there has been limited change in the TFR, but a major change in the timing of fertility toward older ages. Fertility continued to decline until 2001 at ages 15–29, but rose after 1981 at ages 30–34 and older. By 2001 fertility at ages 30–34 had moved marginally ahead of that at ages 25–29, and by 2011 age group 30–34 was clearly Australia's peak childbearing age group. Fertility at this and older ages was not dissimilar to levels that had prevailed at those ages in 1961, but it was the fertility of women who had deferred childbearing into their thirties and forties, not unplanned higher parity fertility attributable to circumscribed ability to avoid such births.

Comparative observations of the type just made on the basis of inspection of Figs. 6.1 and 6.2 can be supplemented by ***simple calculations giving ratios of ASFRs in particular age groups***. Thus, for example, taking ratios of 2011 to 1981 ASFRs in Fig. 6.2 yields values of 0.57, 0.49, 0.71, 1.61, 2.87, 3.38 and 3.00 at ages 15–19 through 45–49. These numbers quantify the extent of the decline in fertility at younger reproductive ages and the increase at older ones over the 30 year period in question as the age pattern of childbearing changed radically whilst the quantum of fertility, as measured by the TFR, barely changed at all.

Measures of the Average Age of Fertility

Differences and changes in age patterns of fertility are also frequently indexed by some sort of measure of what can loosely be termed the ‘average age of fertility’. We have already discussed, in the context of measuring reproductivity, the mean length of generation (MLG). This was defined as ‘the mean age of mothers at the birth of their daughters’, and is an example of one of these measures. There are, however, others, and because standard names and algebraic symbols are not used, the various alternatives are easily confused. Two types of measures should be distinguished: ***measures of the mean age of childbearing***, of which the MLG is an example, and ***measures of the mean age of the fertility schedule (MAFS)*** (a population’s fertility ***schedule*** is its set of ASFRs or ASFFRs). An example of the latter could be calculated from Table 6.1 by multiplying columns (2) and (5) (the age group midpoints and the ASFFR-values), summing the results, and dividing by the sum of column (5). That is:

$$\mathrm{MAFS} = \left[\Sigma_{x=15-19,45-49}\left(b_{(f,x)}/F_x\right)\ \ (m_x)\right] / \,\Sigma_{x=15-19,45-49}\left(b_{(f,x)}/F_x\right) \qquad (6.14)$$

Where all elements mean the same as in Eq. 6.13.

Calculated using the data of Table 6.1, the MAFS has a value of 30.5 years, identical to the value obtained for the MLG, although slightly higher when calculated to more decimal places. These two indices are related to one another in the same way as the NRR and GRR are related. Just as the NRR takes mortality into account and the GRR does not, so, too, the MLG takes mortality into consideration and the MAFS does not. This means that the MLG is invariably lower than the MAFS, because the control for mortality reduces the weights applied in calculation to the midpoints of the reproductive age groups more substantially at the ***older*** reproductive ages (since mortality increases with age). However, in a low mortality population the difference, as here, may be slight.

A value for the MAFS could also be computed using ASFR-values in place of ASFFR-values (i.e., $b_{(f,x)}$ in Eq. 6.14 becomes $b_{(x)}$, meaning ***total*** births to women aged x in year y); if there was no variation in the sex ratio at birth by age of mother the answer would be the same. But the key point is that ***when dealing with ‘average age of fertility’ indices one should be careful to be sure which measure one is dealing with, especially if a comparison is being made***. Since different measures can take on different values for a population, if measures calculated in different ways are compared significance could be read into trends or differences that are wholly or partly spurious.

Two other ‘average age of fertility’ measures that you may encounter are the ***median*** age of the fertility schedule and the ***mean age of mothers giving birth***. The latter has a calculating equation similar in form to Eqs. 6.13 and 6.14, except that the ASFFR-element $b_{(f,x)} / F_x$ is replaced by the number of women giving birth at age x in year y. Once again, be careful to be sure which measure you are dealing with and to only compare like measures.

Natural Fertility and Associated Fertility Models

While graphs of the age pattern of fertility can be instructive, their interpretation in the manner illustrated above lacks precision. Another approach to the description of fertility schedules is to compare them with a ***standard*** schedule representing what is known as ***'natural fertility'***. This term was coined by French demographer Louis Henry, who juxtaposed it with the term ***'controlled fertility'***. Natural fertility refers to ***the fertility of a population which does not consciously practice any fertility restraint***. Henry (1961) distinguished between fertility limiting practices that ***did not*** vary according to a woman's parity and those that ***did*** vary by parity; natural fertility occurred ***where no parity-specific fertility limitation was practiced***. Thus, cultural practices like prolonged breastfeeding, prescribed periods of sexual abstinence and customs regulating the frequency of intercourse, together with other parity-independent fertility inhibitors (e.g., genetic differences affecting fecundity and health conditions influencing pregnancy wastage through miscarriage and stillbirth) are ignored. It follows, of course, that different populations can have different levels of natural fertility, depending on the practices and circumstances that produce ***involuntary*** fertility limitation. 'Controlled fertility' was defined by Henry (1976: 90) as 'the fertility of populations which practice birth control effectively'. Thus defined it is an even less precise concept than 'natural fertility' – 'effectively' presumably means that only planned births occur, but the fertility desires, and hence levels of planned fertility, of populations vary. This is not, however, of major consequence, since it is the ***concept*** of natural fertility that is of analytical importance.

A standard that demographers often use to represent unrestrained, or natural fertility is the ***marital*** fertility schedule achieved by ***Hutterite*** women during 1921–1930. The Hutterites are a small Anabaptist sect living communally in parts of the western plains of the United States and Canada. Their religion encourages large families, they breastfeed for relatively short periods, and they practice no contraception. The observed marital fertility schedule yields a total ***marital*** fertility rate (TMFR) of 14.44 (indicating the number of children the average woman would have if she survived ***and was married*** throughout the reproductive period – essentially it is treated as the TFR attainable by a female population that is universally sexually active throughout this period). The observed schedule, however, includes a marital fertility rate of 0.700 children per woman at ages 15–19. Based on few cases, analysts have deemed this implausibly high and arbitrarily reduced it to 0.300, dropping the TMFR to 12.44. This adjusted Hutterite marital fertility schedule is used in constructing the shortly to be discussed ***Coale fertility indices*** (sometimes also known as the ***Princeton*** fertility indices).

The Coale-Trussell Fertility Model

First, though, we shall consider a way of describing the deviation of a ***marital*** fertility schedule from a natural shape developed by Ansley Coale and James Trussell. This uses another standard schedule of marital fertility which averages ten survey-based schedules described by Henry (1961) as ‘natural’ and which Coale and Trussell (1974) judged to be not significantly affected by age misreporting. The basis of their model is their empirical observation that (1974: 187) ‘marital fertility either follows natural fertility … or departs from natural fertility in a way that increases with age according to a typical pattern.’ The model then specifies that in a population in which there is voluntary control of fertility, the ratio of marital fertility at each age, r(a), to the marital fertility rate from the standard schedule, n(a), is given by:

$$r\,(a)\,/n\,(a) = M\,.\,e^{m\,.\,v(a)} \tag{6.15}$$

Whence:

$$r\,(a) = M\,.\,n\,(a)\ \ .\,e^{m\,.\,v(a)} \tag{6.16}$$

Where r(a) = the observed marital fertility rate for the five-year age group a; n(a) = the marital fertility rate for age group a in the natural fertility schedule; v(a) = an empirically derived value expressing the typical departure from natural fertility at age a due to voluntary fertility control; m = a measure of the extent of fertility control within marriage; M = a scale factor.

The model is applied over the age range 20–24 to 45–49. Age group 15–19 is omitted because, in the words of Coale and Trussell (1974: 188), ‘premarital conceptions have a large and irregular effect on teenage marital fertility.’ The standard schedules of n(a) and v(a) values are given in Table 6.2. The source of the n(a) values was explained above. The v(a) values are also empirically derived, as the means of values obtained for 43 marital fertility schedules published in the 1965 *United Nations Demographic Yearbook*.

Table 6.2 Standard schedules of n(a) and v(a) for the Coale-Trussell fertility model

	Age group					
	20–24	25–29	30–34	35–39	40–44	45–49
n(a)	0.460	0.431	0.395	0.322	0.167	0.024
v(a)	0.000	−0.279	−0.677	−1.042	−1.414	−1.671

Source: Coale and Trussell (1975)

Note: Values in this table differ from those published in Coale and Trussell’s original (1974) paper. They were corrected in an erratum published in *Population Index*, 41(4), 1975

Table 6.3 Calculation of m in the Coale-Trussell Model of Marital Fertility, Philippines 1970

Age group	r(a)	n(a)	r(a)/M.n(a)	$\log_e$ (4)	v(a)	m(a) = (5)/(6)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
20–24	0.327	0.460	1.000	0.000	0.000	
25–29	0.256	0.431	0.835	−0.180	−0.279	0.645
30–34	0.203	0.395	0.723	−0.325	−0.677	0.479
35–39	0.146	0.322	0.638	−0.450	−1.042	0.432
40–44	0.068	0.167	0.573	−0.557	−1.414	0.394
45–49	n.a.	0.024	n.a.	n.a.	−1.671	n.a.
M = r(20 − 24) / n(20 − 24) = 0.711						
m = mean of values in column (7) = 0.488						

The scale factor M is of limited interest since it captures the ***level*** of marital fertility, not its age pattern, which is what we are interested in. It is often estimated as:

$$\mathrm{M} = \mathrm{r}\,(20 - 24)\,/\mathrm{n}\,(20 - 24) \qquad (6.17)$$

The parameter m, measuring the extent of fertility control indicated by comparison of the schedule of r(a) values with the natural fertility schedule of n(a) values, is the value we are most intent on estimating. According to Newell (1988), values of m tend to range from 0.0 to about 2.5. A value of 0.0 implies no fertility control (Coale and Trussell actually suggest that anything below 0.2 is evidence of virtually no control); a value above 2.0 is indicative of extremely strong control. Values of m can be used to group populations according to the extent to which they control fertility within marriage, and can also be used to index trends in marital fertility control over time for individual populations.

With M estimated using Eq. 6.17, values of m (i.e., m(a)) can be found ***for each age group 25–29 to 45–49*** (or 25–29 to 40–44 if the marital fertility rate for age group 45–49 is zero or missing) using the rearranged version of Eq. 6.15 given as Eq. 6.18 below. This rearrangement entails first taking logarithms of both sides of Eq. 6.15. The family of m-values thus obtained ***can then be averaged to yield a single value for the population***. Note that there is no m-value for age group 20–24, the method assuming that marital fertility at those ages is essentially uncontrolled (or as uncontrolled as it gets in the population in question – M, in fact, picks up any deviation from natural fertility at ages 20–24). This is a dubious assumption for many contemporary developed populations, which commonly use highly reliable contraceptive methods to defer childbearing following marriage until economic and career goals have been attained. An example of the calculation of m using this method is presented in Table 6.3.

$$\mathrm{m}\,(\mathrm{a}) = \log_e\left[\mathrm{r}\,(\mathrm{a})\,/\mathrm{M}\,.\,\mathrm{n}\,(\mathrm{a})\right]/\mathrm{v}\,(\mathrm{a}) \qquad (6.18)$$

Table 6.4 Schedule of Hutterite marital fertility rates, 1921–1930

Age group						
15–19	20–24	25–29	30–34	35–39	40–44	45–49
0.300	0.550	0.502	0.447	0.406	0.222	0.061

Although frequently used, the approaches to evaluating M and m just described have some limitations. The main one is that M is obtained with reference to only one age group, and thereafter influences all subsequent calculations. An alternative approach again entails first taking logarithms of both sides of Eq. 6.15 to yield the equation:

$$\log_e [r(a)/n(a)] = \log_e M + m \,.\, v(a) \tag{6.19}$$

If we let $y = \log_e [r(a)/n(a)]$, $c = \log_e M$, and $x = v(a)$, Eq. 6.19 has the form $y = mx + c$, which is the equation of a straight line. This can be fitted to pairs of (x,y) values for the various five-year age groups (which we can obtain since r(a), n(a) and v(a) are all known) using ordinary least squares regression to yield values of m and c, and hence values of m and M. Applied to the example in Table 6.3, this approach gives estimates of $m = 0.379$ and $M = 0.682$. Comparing the former value with that computed in Table 6.3 a somewhat lower level of marital fertility control is suggested, and it is clear that in using the Coale-Trussell model it is unwise to use different methods of estimating m-values which are to be compared.

The Coale (or Princeton) Fertility Indices

Ansley Coale's indices I_f (overall fertility), I_g (marital fertility), I_h ('illegitimate', or unmarried, fertility) and I_m (marriage index) were developed in the context of a major study of the decline of fertility in Europe during the nineteenth and twentieth centuries conducted at Princeton University from the late 1960s. Data on annual births generally were not available by age of mother, so that TFRs could not be calculated (GFRs could be, but did not control for age). Coale's solution was to develop three fertility indices, ***I_f, I_g and I_h***, which are in fact indirectly standardized ratios (see Chap. 2). They ***express the levels of fertility (total, marital and 'illegitimate' (non-marital)) in a population as proportions of natural fertility as represented by the previously mentioned 1921–1930 Hutterite marital fertility schedule*** (adjusted at ages 15–19). The method simply applies the Hutterite schedule to distributions of total women, married women and unmarried women by five-year age groups to yield ***expected*** numbers of births to each group assuming natural fertility prevailed, then calculates ratios of observed to expected numbers of births. The Hutterite marital fertility schedule is given in Table 6.4.

Calculation of the Coale indices requires two types of data: (i) total births in a year classified by whether or not the mother was married (births may also be averaged over a short series of consecutive years), and (ii) the distribution of the female population at mid-year (or the middle of the period over which annual births are averaged) by marital status and five-year age groups. Calculating equations for the three fertility indices are as follows.

Index of overall fertility:

$$I_f = B_t / \Sigma_{i=15-19,45-49} W_i \ . \ H_i \qquad (6.20)$$

Where B_t = total births in year y (or the average annual number of births over a short period centred on year y); W_i = the mean (or mid-year) total number of women aged i in year y; H_i = the marital fertility rate of Hutterite women aged i.

Index of marital fertility:

$$I_g = B_n / \Sigma_{i=15-19,45-49} M_i \ . \ H_i \qquad (6.21)$$

Where B_n = nuptial (marital) births in year y (or the average annual number of such births over a short period centred on year y); M_i = the mean (or mid-year) number of married women aged i in year y; H_i = the marital fertility rate of Hutterite women aged i.

Index of unmarried fertility:

$$I_h = B_e / \Sigma_{i=15-19,45-49} U_i \ . \ H_i \qquad (6.22)$$

Where B_e = ex-nuptial (non-marital) births in year y (or the average annual number of such births over a short period centred on year y); U_i = the mean (or mid-year) number of unmarried women aged i in year y; H_i = the marital fertility rate of Hutterite women aged i.

There was, of course, a fourth index listed at the beginning of this section. This is ***I_m*, *the marriage index***, or index of the proportion married among women in the reproductive age group. It is treated separately because it is not a fertility index. Given, however, that in virtually all societies married women are much more likely to bear children than are unmarried women, clearly a potentially major reason for differences/changes in overall fertility is differences/ changes in the extent to which women of reproductive age are married. I_m is designed to recognize this reality and to allow it to be taken into account in interpreting fertility differentials and trends. It can be calculated as either a weighted average of the proportions of women married in five-year age groups 15–19 to 45–49 years, with the Hutterite marital fertility schedule providing the weights, or as the ratio of expected births to married women to expected births to total women assuming Hutterite marital fertility. Thus, either:

$$I_m = \Sigma_{i=15-19,45-49} (M_i / W_i) \ . \ H_i / \Sigma_{i=15-19,45-49} H_i \qquad (6.23)$$

Where M_i = the mean (or mid-year) number of married women aged i in year y; W_i = the mean (or mid-year) total number of women aged i in year y; H_i = the marital fertility rate of Hutterite women aged i.

Table 6.5 Calculation of Coale fertility indices I_f, I_g, I_h and I_m for Bulgaria, 1985

Age group	M_i	U_i	W_i	H_i	$W_i \,.\, H_i$	$M_i \,.\, H_i$	$U_i \,.\, H_i$
15–19	40,168	259,481	299,649	0.300	89,895	12,050	77,844
20–24	195,672	98,129	293,801	0.550	161,591	107,620	53,971
25–29	259,417	50,178	309,595	0.502	155,417	130,227	25,189
30–34	278,240	41,916	320,156	0.447	143,110	124,373	18,736
35–39	297,120	41,258	338,378	0.406	137,381	120,631	16,751
40–44	246,680	34,084	280,764	0.222	62,330	54,763	7,567
45–49	236,105	36,189	272,294	0.061	16,610	14,402	2,208
					$\sum = 766{,}334$	$\sum = 564{,}066$	$\sum = 202{,}266$
$B_t = 118{,}955$; $B_n = 105{,}001$; $B_e = 13{,}954$							
$I_f = 118{,}955 / 766{,}334 = 0.155$							
$I_g = 105{,}001 / 564{,}066 = 0.186$							
$I_h = 13{,}954 / 202{,}266 = 0.069$							
$I_m = 564{,}066 / 766{,}334 = 0.736$							

Source: United Nations demographic yearbook, 1986 and 1990

Or:

$$I_m = \Sigma_{i=15-19,45-49} M_i \,.\, H_i / \Sigma_{i=15-19,45-49} W_i \,.\, H_i \tag{6.24}$$

Where all elements mean the same as in Eq. 6.23.

Equation 6.24 is more convenient because the two summations it requires are also required in evaluating I_f (Eq. 6.20) and I_g (Eq. 6.21), and it is a simple matter to divide one by the other. An example of the calculation of I_f, I_g, I_h and I_m is presented in Table 6.5.

An important relationship between the Coale indices is expressed by the equation:

$$I_f = I_m \,.\, I_g + (1 - I_m) \,.\, I_h \tag{6.25}$$

And if I_h (unmarried fertility) is negligible, as it often is and/or historically has been in many populations, this reduces to the approximate relationship:

$$I_f = I_m \,.\, I_g \tag{6.26}$$

Thus in populations where non-marital fertility is very low (not more than, say, 3–4%) we have an index of overall fertility that is approximately the product of an index of marital fertility and an index of the proportion married.

To illustrate the use of the Coale indices, Fig. 6.3 plots values for Australia for census years between 1861 and 2011. In 1861 overall fertility (I_f) was about 50 % of the Hutterite natural level, while marital fertility (I_g) was around 70 % of it. Overall fertility fell between 1861 and 1881, but marital fertility rose to peak at 74 % of

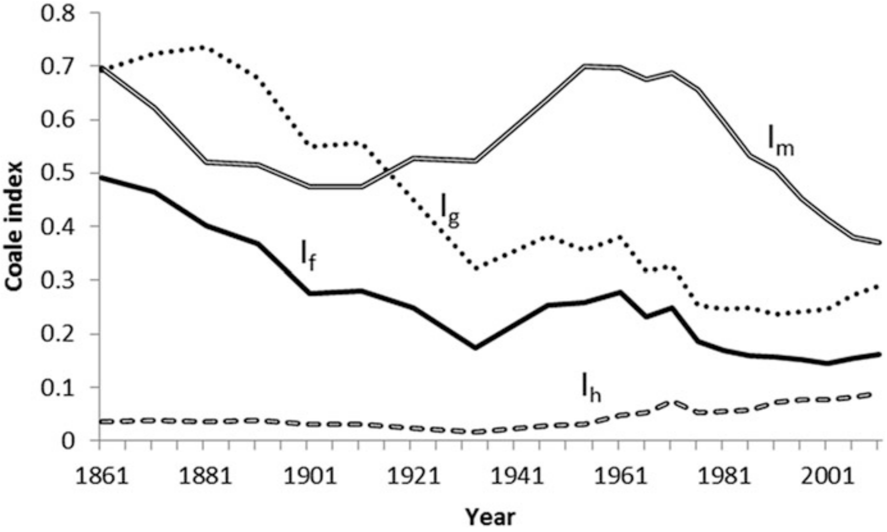


Fig. 6.3 Coale fertility indices for Australia, census years 1861 to 2011 (Source: Jones (1971), author's calculations)

the Hutterite level. Thus, overall fertility declined solely because the marriage index (I_m) dropped sharply. Women married later, thereby delaying the commencement of childbearing, but did not reduce the regularity with which they had children once married.

The period between the censuses of 1881 and 1933 saw overall fertility continue to decline, reaching a level by the latter date that was less than 18 % of the Hutterite level. Over this period the driving force behind overall fertility decline clearly was decline in marital fertility, which more than halved from 74 % of the Hutterite level to 32 %. This was the period of Australia's fertility transition, the decline in marital fertility accelerating between 1881 and 1901, easing off in the first decade of the twentieth century as depression conditions of the 1890s eased, then reasserting itself between 1911 and the Great Depression of the early 1930s. During this period the marriage index remained quite stable, a slight rise between 1911 and 1921 partly offsetting the decline in marital fertility and moderating the decline in overall fertility.

Between 1933 and 1954 the marriage index rose rapidly. Indeed, had there been a census in the late 1930s an even steeper increase would be apparent, detailed studies showing this 'marriage boom' to have really commenced with the outbreak of the Second World War. This trend reinforced a modest resurgence in marital fertility between 1933 and 1947, and then more than offset a slight decline between 1947 and 1954, yielding an upward trend in overall fertility that continued until 1961. Subsequently, with oral contraception introduced in Australia that year, the trends in both overall and marital fertility were generally downward until marital fertility picked up again slowly through the 1990s and more quickly after 2001. The marriage index also fell sharply again after 1971 as trends to younger and more universal marriage reversed emphatically with the rise to prominence of consensual partnering, and this drove overall fertility down more rapidly than marital fertility. The period since 1951 has also seen unmarried fertility increase in importance as, first, younger women and men became more premaritally sexually active, and then as they gained some control over unplanned fertility resulting from this trend, the practice of couples cohabiting and having children in such relationships became more widespread. The brief break in this trend between 1971 and 1976 marks the transition from non-marital fertility rising largely as a result of unintended teenage and early adult pregnancies generated by the greater sexual freedom of the 1950s and 1960s, to it rising at more normative reproductive ages due to the spread of cohabitation and childbearing within it. That five-year period brought a massive improvement in unmarried teenage and early adult fertility control as access was finally gained to the pill (unmarried women had largely been denied it through the 1960s) and induced abortion became readily available after a landmark legal case in New South Wales gave the green light to stand alone abortion clinics.

Other Fertility Models

Aside from fertility models based on notions of natural fertility, various other types have been proposed. Two will be dealt with here: William Brass's ***relational Gompertz model*** and John Bongaarts's model relating the more important of what are known as the ***proximate determinants of fertility*** to the observed fertility level of a population.

The Brass Relational Gompertz Model

This model aims to describe any schedule of age-specific fertility rates by two parameters, α and β, which capture the difference between that fertility schedule ***expressed in cumulative proportional form*** and a standard fertility schedule expressed in similar form. To express a fertility schedule in cumulative proportional form one cumulates ASFRs for single-year or five-year age groups successively from the youngest to the oldest reproductive age (multiplying the latter by 5 in recognition that a woman takes five years to pass through any five-year age group), assigns the result after each new ASFR is added to the exact age defining the upper bound of the age group in question, and divides each of these ***cumulative fertility rates*** by the TFR. The outcome is a schedule of cumulative fertility to successive exact ages expressed as a proportion of ultimate completed fertility (i.e., values rise from zero at the exact age marking the lower bound of the youngest age group in which births occur to 1.0 at that marking the upper bound of the oldest age group in which they occur).

The standard fertility schedule used in this model was designed to reflect the typical pattern of fertility in high fertility populations. Table 6.6 shows both cumulative proportional and Gompertz-transformed values from this schedule for one-year intervals of exact age. For use with five-year age group data simply extract Gompertz-transformed values for exact ages 20, 25, 30, 35, 40 and 45. The basis of the Brass relational Gompertz approach is the empirical observation that applying a Gompertz transformation to any fertility schedule expressed in cumulative proportional form yields approximately a straight line when plotted against exact age, and thus approximately a straight line when plotted against any other, or a standard, Gompertz-transformed fertility schedule as well. The method focuses on the straight line resulting from the plot of Gompertz-transformed observed and standard fertility schedules against one another. The parameters α and β are the y-intercept and slope, respectively, of this straight line, which has the equation:

$$G_x = \alpha + \beta G_{s,x} \tag{6.27}$$

Where G_x = the Gompertz-transformed ***observed*** cumulative proportional fertility rate to exact age x; $G_{s,x}$ = the Gompertz-transformed ***standard*** cumulative proportional fertility rate to exact age x.

If the observed and standard fertility schedules were identical the straight line represented by Eq. 6.27 obviously would intercept at the origin and have a slope of 1.0; hence $\alpha = 0$ and $\beta = 1.0$ are the ‘no difference’ benchmarks against which computed values of α and β are interpreted. According to Newell (1988: 177–178), α ‘changes the age-location of the model’ while β ‘may be interpreted as determining the spread, or degree of concentration, of the schedule.’ A value of $\alpha = 0$ indicates that, as in the standard schedule, about half of total childbearing occurs by age 27. Positive values of α indicate earlier fertility than the standard; negative values indicate later fertility. Values of +0.5 and −0.5 respectively suggest ***very*** early and ***very*** late age patterns of fertility. Values of β lower than 1.0 (down to perhaps 0.65) indicate a wider spread of fertility by age than in the standard population (as, in the limiting case, might occur in natural fertility populations which also marry early), while values higher than 1.0 indicate a narrower spread of fertility (values above 1.5 suggest a late start, and early finish, to childbearing, as occurs in some contemporary European populations).

Having estimated α and β in a manner described below it is possible to work backwards from Eq. 6.27 and ‘fit’ ASFR values that would ***exactly*** satisfy the equation for each of the standard $G_{s,x}$ values (remember that we referred above to the Gompertz-transformed observed and standard cumulative proportional fertility schedules ***approximating*** a straight line relationship). If these ‘fitted’ values were judged, in comparison with the original ASFRs, to be a ‘good’ fit, one practical application of Eq. 6.27 would be the estimation of a schedule of ***single***-year ASFRs from the observed ***five***-year values. Having used the latter to specify the relevant form of Eq. 6.27, one would work backwards from it, varying x in single-

Table 6.6 Standard cumulative proportional and Gompertz-transformed fertility rates for the Brass relational Gompertz fertility model

Exact age	Cumulative proportional fertility rate	Gompertz-transformed rate	Exact age	Cumulative proportional fertility rate	Gompertz-transformed rate
11	0.00000	3.18852	31	0.65016	−0.84272
12	0.00000	2.70008	32	0.68968	−0.99014
13	0.00002	2.37295	33	0.72722	−1.14407
14	0.00035	2.07262	34	0.76275	−1.30627
15	0.00277	1.77306	35	0.79618	−1.47872
16	0.01168	1.49286	36	0.82751	−1.66426
17	0.03043	1.25061	37	0.85663	−1.86597
18	0.05826	1.04479	38	0.88354	−2.08894
19	0.09428	0.85927	39	0.90816	−2.33192
20	0.13584	0.69130	40	0.93019	−2.62602
21	0.18187	0.53325	41	0.94925	−2.95500
22	0.22993	0.38524	42	0.96480	−3.32873
23	0.27897	0.24423	43	0.97698	−3.75984
24	0.32829	0.10783	44	0.98591	−4.25499
25	0.37731	−0.02564	45	0.99188	−4.80970
26	0.42597	−0.15853	46	0.99555	−5.41311
27	0.47371	−0.29147	47	0.99782	−6.12864
28	0.52013	−0.42515	48	0.99915	−7.07022
29	0.56517	−0.56101	49	0.99982	−8.64839
30	0.60861	−0.70000	50	1.00000	+∞

Table 6.7 Brass relational Gompertz fertility model fitted to data for Ecuador, 1993

						Fitted	Fitted	Fitted	Fitted
Age group	ASFR	F_x	F_x / TFR	G_x	$G_{s,x}$	G_x ($G_{f,x}$)	F_x / TFR	F_x	ASFR
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
15–19	109.4	0.5470	0.1511	0.6365	0.6913	0.6026	0.1609	0.5823	116.5
20–24	191.6	1.5050	0.4158	−0.1306	−0.0256	−0.1424	0.4201	1.5206	187.6
25–29	178.9	2.3995	0.6629	−0.8888	−0.7000	−0.8431	0.6503	2.3538	166.6
30–34	133.1	3.0650	0.8468	−1.7940	−1.4787	−1.6523	0.8256	2.9883	126.9
35–39	77.9	3.4545	0.9544	−3.0646	−2.6260	−2.8445	0.9435	3.4150	85.3
40–44	26.8	3.5885	0.9914	−4.7517	−4.8097	−5.1136	0.9940	3.5978	36.6
45–49	6.2	3.6195	1.0000				1.0000	3.6195	4.3
$\alpha = -0.1158$; $\beta = 1.0391$									

year intervals and substituting single-year-of-age Gompertz-transformed cumulative proportional fertility rates for the standard population from Table 6.6.

An example should now help to bring the Brass relational Gompertz procedure to life. In column (2) of Table 6.7 are presented five-year ASFRs for Ecuador in 1993. The subsequent calculations involve the following steps.

Step 1: In column (3), x is the exact age marking the upper bound of each age group (i.e., 20, 25, …, 50) and F_x is the cumulative fertility rate ***per woman*** to exact age x. Thus:

$$F_{20} = 5\ .\ \mathrm{ASFR}_{15-19}/1,000$$
$$F_{25} = 5\ .\ (\mathrm{ASFR}_{15-19} + \mathrm{ASFR}_{20-24})\,/1,000$$
$$F_{30} = 5\ .\ (\mathrm{ASFR}_{15-19} + \mathrm{ASFR}_{20-24} + \mathrm{ASFR}_{25-29})\,/1,000$$
$$\ldots\ldots\ldots\ldots$$
$$F_{50} = 5\ .\ (\mathrm{ASFR}_{15-19} + \cdots + \mathrm{ASFR}_{45-49})\,/1,000$$

Step 2: Column (4) is obtained by dividing each value in column (3) by the TFR (i.e., by F_{50}) to yield the cumulative ***proportional*** fertility rate to exact age x.

Step 3: In column (5) the cumulative proportional fertility rates of column (4) are Gompertz-transformed using the equation:

$$G_x = \log_e(-\log_e(F_x/\mathrm{TFR})) \tag{6.28}$$

Step 4: Column (6) gives Gompertz-transformed cumulative proportional fertility rates for the standard fertility schedule extracted from Table 6.6.

Step 5: We now need to solve for α and β in Eq. 6.27 using the values of G_x and $G_{s,x}$ in Table 6.7. One way to do so would be by ordinary least squares regression, but a simpler approach commonly is used. The six pairs of G_x, $G_{s,x}$ values in Table 6.7 are divided into two groups of three (x = 20, 25 and 30; x = 35, 40 and 45), and for each group the three equations of the form of Eq. 6.27 created by substituting G_x, $G_{s,x}$ pairs are summed to yield two simultaneous equations in α and β. Thus:

$$G_{20} + G_{25} + G_{30} = (\alpha + G_{s,20}\ .\ \beta) + (\alpha + G_{s,25}\ .\ \beta) + (\alpha + G_{s,30}\ .\ \beta)$$
$$\text{i.e.,}\quad 0.6365 - 0.1306 - 0.8888 = 3\alpha + (0.6913 - 0.0256 - 0.7000)\,\beta$$
$$\text{or} - 0.3829 = 3\alpha - 0.0343\beta$$

$$\text{And}:\quad G_{35} + G_{40} + G_{45} = (\alpha + G_{s,35}\ .\ \beta) + (\alpha + G_{s,40}\ .\ \beta) + (\alpha + G_{s,45}\ .\ \beta)$$
$$\text{i.e.,} -1.7940 - 3.0646 - 4.7517 = 3\alpha + (-1.4787 - 2.6260 - 4.8097)\,\beta$$
$$\text{or} - 9.6103 = 3\alpha - 8.9144\beta$$

Solving these two simultaneous equations gives α = –0.1158 and β = 1.0391.

Step 6: To obtain the fitted values of G_x shown in column (7) of Table 6.7 we simply apply Eq. 6.27 treating G_x as the unknown (call it $G_{f,x}$ if you like, where f stands for 'fitted') and substituting the values of α and β just calculated together with the relevant value of $G_{s,x}$. The result is a set of G_x values that exactly satisfy for each value of $G_{s,x}$ the straight line whose equation we found by estimating α and β at step 5.

Step 7: The fitted values of G_x are converted to fitted ASFRs by reversing the procedure that was followed in generating column (5) from column (2). We first obtain, at column (8), fitted values of F_x / TFR (i.e., fitted cumulative proportional fertility rates to exact age x) by ‘undoing’ the Gompertz transformations of the fitted G_x-values. To do this use:

$$\text{fitted}\,(F_x/\text{TFR}) = \text{exponent}\,(-\text{exponent}\ (\text{fitted}\ G_x)) \tag{6.29}$$

Next, calculate fitted F_x values (column (9)) by multiplying the fitted F_x/TFR values by the original TFR (i.e., by F_{50} from column (3)). You can perhaps appreciate at this point that the fitting procedure is effectively redistributing recorded total fertility among the reproductive age groups.

The fitted F_x values give cumulative (total) fertility to exact ages 20, 25, …, 50. For each five-year age group (except 15–19), find the fitted portion of total fertility associated with the age group by subtracting the fitted F_x associated with its lower bound (remember x is an exact age) from that associated with its upper bound (e.g., for age group 20–24 find fitted F_{25} - fitted F_{20}) (the fitted portion for age group 15–19 is simply fitted F_{20}). Then divide each fitted portion by 5 and multiply by 1,000 to yield the fitted ASFRs of column (10). Remember, we did the reverse of this (multiplied by 5 and divided by 1,000) to our original ASFRs at step 1.

The parameters α and β estimated in this example suggest that fertility in Ecuador in 1993 occurred a little later than in the standard population and tended to be marginally more compressed in terms of the ages over which it was spread. The match between the fitted ASFRs and the original ASFRs is not especially good, and so we would probably hesitate to use the particular form of Eq. 6.27 obtained using our estimates of α and β to generate single-year-of-age ASFRs.

The Bongaarts Fertility Model

In 1956 the American demographers Kingsley Davis and Judith Blake published a ground-breaking paper in which they proposed that social, economic and cultural conditions affected a population’s fertility through a set of 11 ***intermediate fertility variables***. These were the variables which ***directly influenced*** fertility, and through which any other factor (e.g., the level of female education) had to operate. Davis and Blake divided these intermediate variables into three subgroups: ***Factors affecting exposure to intercourse*** (age of entry into sexual unions, extent of permanent female celibacy, amount of the reproductive period lost to marital disruption or widowhood, extent of voluntary abstinence, extent of involuntary abstinence, and coital frequency when not abstaining); ***factors affecting exposure to conception*** (natural forces affecting fecundity, extent of contraception, and voluntary practices affecting fecundity); and ***factors affecting gestation and successful parturition (childbirth)*** (foetal mortality from involuntary causes, and foetal mortality from voluntary causes).

The Davis-Blake framework found wide acceptance in broad concept, but in its specifics proved difficult to incorporate into quantitative models. For quite some time those models that were developed were extremely complex. John Bongaarts's contribution has been to greatly simplify the modelling of fertility in terms of the intermediate fertility variables or, as they have come more commonly to be known these days, the ***proximate determinants of fertility***.

The basis of this simplification is (i) a respecification of the proximate determinants to a list of seven (the female marriage pattern, the extent and effectiveness of contraception, the prevalence of induced abortion, the duration of post-partum infecundablility, natural fecundability (indexed by frequency of intercourse), the prevalence of spontaneous intrauterine mortality (miscarriage), and the extent of permanent sterility), and (ii) a further collapsing of this list to four 'principal proximate determinants' by eliminating the last three as not sufficiently variable from population to population to be important in accounting for fertility differentials and trends. Thus, in the Bongaarts fertility model the four principal proximate determinants are considered the prime potential inhibitors of fertility to below its theoretical maximum level – delayed or prematurely ended marriage, contraception, induced abortion, and post-partum infecundability induced by breastfeeding or abstinence.

The model proposes four levels of fertility from which, if known, impacts of the four principal proximate determinants could be derived. With the inhibiting effects of all four determinants operative a population's fertility level is measured by its TFR (based on age-specific ratios of ***marital*** births ***only*** to total women). If the fertility-inhibiting effect of delayed and disrupted marriage is eliminated, fertility rises to a level TM, the total ***marital*** fertility rate (based on age-specific ratios of marital births to ***married*** women). If the effects of contraception and induced abortion are then also eliminated it rises further to a level TN, the total ***natural*** marital fertility rate. And finally, if breastfeeding and post-partum abstinence are non-existent it rises to a level TF, the total ***fecundity*** rate. This last rate is set by the model at 15.3 children per woman. TFs vary from population to population because of the effects of the three proximate determinants disregarded by the model, but do so over a fairly narrow range. The value 15.3 is an average of this narrow range of empirical values.

Against this theoretical background Bongaarts defines four indices which measure the levels of the principal proximate determinants on scales from 0 to 1, where 0 corresponds with maximum inhibition of fertility and 1 with zero inhibition. They are:

C_m = index of marriage (the proportion of women of reproductive age married) = 1 where marriage is universal and 0 where there is no marriage.

C_c = index of contraception = 1 where contraception is totally absent and 0 where all fecund women use 100 % effective contraception.

C_a = index of induced abortion = 1 where there is no induced abortion and 0 where all pregnancies are aborted.

C_i = index of post-partum infecundability = 1 where there is no breastfeeding or post-partum abstinence and 0 where the duration of infecundability from these practices is infinite.

These indices can, by definition, be expressed in terms of the four fertility levels as follows:

$$C_m = TFR/TM \tag{6.30}$$

$$C_c \,.\, C_a = TM/TN \tag{6.31}$$

$$C_i = TN/TF \tag{6.32}$$

From these equations it follows that:

$$TFR = C_m \,.\, (C_c \,.\, C_a) \,.\, C_i \,.\, TF \tag{6.33}$$

Or

$$TFR = C_m \,.\, (C_c \,.\, C_a) \,.\, C_i \,.\, 15.3 \tag{6.34}$$

Equation 6.34 is the one that specifies the basic Bongaarts model. Obviously C_m, $(C_c \,.\, C_a)$ and C_i can be calculated from Eqs. 6.32, 6.33, and 6.34 if the fertility rates TFR, TM, TN and TF are known, but it is rare for all to be known (TN and population-specific values of TF, in particular, are hard to obtain). Thus Bongaarts provides a series of alternative equations for estimating his indexes of the principal proximate determinants.

The index of marriage is given by:

$$C_m = \Sigma_{x=15-19,45-49}\ m_x \,.\, g_x / \Sigma_{x=15-19,45-49}\ g_x \tag{6.35}$$

Where m_x = the age-specific proportion of women aged x currently married (or in a consensual union); g_x = the marital fertility rate at age x.

Equation 6.35 indicates C_m to be a weighted average of age-specific proportions of women currently married (or in a consensual union), with age-specific marital fertility rates providing the weights. It is not merely the proportion of all women of reproductive age who are married (or in a consensual union), because the fertility impact of marriage depends partly on the age distribution of married women (marital fertility rates varying with age).

If values of g_x are not directly available they can be estimated as $g_x = f_x / m_x$, where f_x is the age-specific fertility rate at age x based on marital births only. The frequent parenthetic reference to 'or in a consensual union' is indicative that measurement problems can arise for populations where such unions are sufficiently common that they cannot be ignored. Note in particular that births to consensually partnered women should be regarded as ‘marital’ births, but they are frequently tabulated as, and impossible to distinguish from among, ***non***-marital births. Bongaarts also highlights possible difficulty in estimating g_{15-19} satisfactorily using the relationship $g_x = f_x / m_x$, advocating that in this circumstance a reasonable approximation is to set $g_{15-19} = 0.75\ g_{20-24}$.

The index of contraception is given by:

$$C_c = 1 - 1.08\ .\ u\ .\ e \tag{6.36}$$

Where u = the proportion of married women of reproductive age currently using contraception (including male methods, sterilization and abstinence other than post-partum); e = the average use-effectiveness of the methods being used.

Estimates of u mostly are obtained from fertility or contraceptive prevalence surveys. Values of e are computed from data on methods of contraception used and standard method-specific values of use-effectiveness using:

$$e = \Sigma e_m\ .\ u_m/u \tag{6.37}$$

Where e_m = the use-effectiveness level for method m; u_m = the proportion of married women of reproductive age using method m; $u = \sum u_m$.

When writing in the early 1980s, Bongaarts had ‘reasonably reliable’ method-specific use-effectiveness data only for the Philippines (sterilization = 1.00, IUD = 0.95, pill = 0.90, other methods = 0.70) and the United States (sterilization = 1.00, pill = 0.99, IUD = 0.97, condom = 0.94, diaphragm = 0.92, foams, creams, jellies = 0.91, rhythm = 0.87, other = 0.93). Obviously population-specific schedules of use-effectiveness are the ideal, but in their absence one’s options are to select the most plausible schedule for another population available or to decide that the evaluation of Eq. 6.36 cannot reliably be completed.

The coefficient 1.08 in Eq. 6.36 is a correction factor for the reality that couples do not use contraception if they know or believe their unions to be involuntarily sterile. This empirically-based coefficient in effect inflates u to yield the proportion of married women of reproductive age ***in non-sterile unions*** currently contracepting. The idea is that what is needed is a measure of contraceptive practice ***in situations where it has fertility-inhibiting potential***.

Bongaarts’s equation for his index of induced abortion is empirically based and designed to reflect research findings (i) that on average an induced abortion averts less than one birth (because miscarriage or stillbirth might independently have prevented a live birth, and because a woman resumes ovulation and hence exposure to the risk of a further pregnancy sooner after induced abortion than if she has a live birth), and (ii) that the number of births averted per induced abortion varies depending on the extent to which contraception is practiced following the induced abortion. The equation is:

$$C_a = TFR/\left[TFR + 0.4\ .\ (1+u)\ .\ TA\right] \tag{6.38}$$

Where TFR = total fertility rate (again based on marital births only); u means the same as in Eq. 6.36; TA = the total abortion rate (the sum over the reproductive ages of single-year age-specific abortion rates, or five times the sum of five-year age-specific rates, ***excluding abortions to unmarried women***).

This equation is the ratio of the observed TFR to the estimated TFR in the absence of induced abortion. If induced abortion is so infrequent as to be virtually non-existent, $C_a = 1.0$.

Finally, the index of post-partum infecundability is given by:

$$C_i = 20/\left(18.5 + i\right) \tag{6.39}$$

Where i = the average duration of post-partum infecundability caused by breastfeeding or post-partum abstinence in months.

Values of i are generally not readily available, but can be estimated from survey data on the duration of breastfeeding using:

$$i = 1.753\ .\ e^{0.1396B - 0.001872B^2} \tag{6.40}$$

Where B = mean duration of breastfeeding in months.

The Bongaarts model can, of course, be used to estimate the value of any of the principal proximate determinant indices given values of the other indices, and use can also be made of the equations linking the indices to the fertility levels TFR, TM, TN and TR to evaluate individual indices or fertility levels. It can be applied most straightforwardly to populations in which sexual activity is more or less confined to marriage (whence a ‘regular’ TFR will reflect only marital births) and consensual partnering is virtually non-existent whence there are no complications in distinguishing ‘married’ women and ‘marital’ (births). Where these conditions are not met particular care is needed to ensure that variables are measured appropriately, and the scope for meeting insurmountable measurement problems increases.

There are also a variety of applications of the Bongaarts model besides the simple calculation of individual indices and fertility levels. Bongaarts and Potter (1983) discuss several. While data are rarely if ever available to enable fertility transitions for individual populations to be traced, those for a range of populations recently at different stages of transition have been used to construct a ***synthetic*** profile of the ‘typical’ transition from natural to controlled fertility. This suggests that transition is associated with a shortening of post-partum infecundability, a large rise in contraception, and a decline in the proportion married. Actual transitions for contemporary developing populations are claimed to closely resemble this pattern, but historical European transitions to differ in that proportions of women married ***rose*** and mean ages at marriage declined. Other applications discussed are the decomposition of change in fertility (to assess the proximate determinants most responsible), assessment of the proximate determinants underpinning socioeconomic differences in marital fertility, the construction of models linking ***age-specific*** fertility rates to the proximate determinants, detailed analysis of an empirical relationship between the CBR and contraceptive prevalence, projection of the fertility consequences of a nominated change in contraceptive practice, and assessment of the contraceptive prevalence level required to meet a fertility target.

We will conclude this section by looking briefly at the penultimate of these applications. Suppose that we are interested in ***the potential effect of a family planning program***. If TFR_1 refers to the total fertility rate before the program and TFR_2 to its value at a later point after the program has had a chance to operate, the proportionate change in the TFR can be written:

$$(TFR_1 - TFR_2)\,/TFR_1 = 1 - (TFR_2/TFR_1)$$

Which in terms of the Bongaarts model (Eq. 6.34) is:

$$\begin{aligned}&(TFR_1 - TFR_2)\,/TFR_1\\&= 1 - (C_{m,2}\,.\,C_{c,2}\,.\,C_{a,2}\,.\,C_{i,2}\,.\,15.3)\,/\,(C_{m,1}\,.\,C_{c,1}\,.\,C_{a,1}\,.\,C_{i,1}\,.\,15.3)\end{aligned}$$

Now if it is reasonable to assume that the female marriage pattern, the abortion level among married women and the average period of post-partum infecundability remain constant over the program period, this reduces to:

$$(TFR_1 - TFR_2)\,/TFR_1 = 1 - C_{c,2}/C_{c,1}$$

Substituting using Eq. 6.36 and simplifying then yields:

$$TFR_2 = TFR_1\,(1 - 1.08\,.\,u_2\,.\,e_2)\,/\,(1 - 1.08\,.\,u_1\,.\,e_1) \qquad (6.41)$$

By substituting the pre-program TFR, known pre-program contraceptive use (u_1) and use-effectiveness (e_1) values, and ***target post***-program use (u_2) and effectiveness (e_2) values it is possible to estimate a post-program TFR, and hence the reduction in TFR achievable if target use and effectiveness levels are reached. It might be that one would wish to assume no change in use-effectiveness, in which case $e_2 = e_1$, and the assessment that is made is an assessment of the fertility reduction achievable by increasing contraceptive use to level u_2.

This type of practical application to which the Bongaarts fertility model lends itself makes it especially valuable. It is appropriate to end this discussion of fertility models, however, with a warning that ***all*** fertility models should be used very cautiously for purposes other than summary description. They ***are*** only ***models***, and placing ***blind*** faith in them can be perilous.

Analysis of Birth Intervals

Generally speaking, a birth interval is ***the length of time between successive live births***, and is usually measured in months. Stillbirths and abortions are ignored. Two birth intervals, however, do not conform to this definition. The ***first birth interval***, or ***protogenetic*** interval, is the interval from marriage to first birth. It may be negative if the first birth precedes marriage, and non-existent for a woman who has had one or more children without ever marrying (she may, though, have second, third, etc. birth intervals). Some populations even record negative ***average*** first birth intervals, but there is really not much point to analysis of first birth intervals in such circumstances. The concept of the first birth interval presupposes that marriage marks the commencement of exposure to the risk of conception, or at least that fertility, by cultural convention, takes place within marriage (so that if first conception precedes marriage there is strong social pressure to marry before the child is born). Where these conditions more or less prevail, change in the average length of the first birth interval for successive cohorts of women can provide useful information about change in fertility within marriage, and more particularly about change in the onset of the entire process of family formation. However, where large proportions of first births occur to unmarried women, or even where premarital cohabitation is common and marriage then is widely deliberately timed to coincide with the transition to parenthood, some other reference point than marriage (first coitus or entry into a first union) seems more relevant. A more general definition of the first birth interval as the interval from first exposure to the risk of conception (defined in one of the ways just suggested) to the first live birth sometimes is encountered, but unless a source clearly states that it has been used you should assume first marriage as the reference point.

The second birth interval not conforming to the general definition is the ***open birth interval***. This is the interval, again in months, from the most recent birth, or from marriage in the case of childless married women, to the date of data collection. Birth intervals which ***do*** conform to the general definition are labelled according to the order of the birth marking the ***end*** of the interval; thus the interval from the first to the second birth is the second birth interval, etc. These birth intervals are known as the ***intergenetic*** intervals. Like the protogenetic interval they are ***closed intervals*** – closed by the birth that ended the interval.

You will also encounter in the demographic literature various other intervals closely related to birth intervals as just defined. Sometimes both live births and stillbirths (i.e., total births) are taken into account. There can in particular be some justification for this in populations with high levels of infant mortality, because differentiating stillbirths from very early neonatal deaths may be both problematic and somewhat artificial. More generally it can be argued that having a stillbirth indicates intent to have a live birth, and that explaining birth intervals that have been lengthened by the unforeseen occurrence of a stillbirth is difficult (although miscarriages raise similar issues). Other intervals that may be focused on are the ***interval from first exposure to first conception***, ***intervals from one live birth to***

conception of the next, and ***inter-pregnancy intervals***. The last-mentioned can be (i) the interval from the end of one pregnancy (however terminated) to the beginning of the next (again, regardless of whether it results in a live birth), or (ii) the interval between the events (whatever they may be) marking the end of successive pregnancies. Clearly one should be certain to check in any practical situation which definition has been used. But these three types of intervals tend to be of interest when the aim is to measure exposure to the risk of ***conception*** when studying issues such as contraceptive effectiveness. We will not, however, concern ourselves further with them here.

Data on birth and related intervals generally derive from ***pregnancy or birth histories***, which date key events in women's reproductive careers. Intervals are computed by subtracting the date marking the beginning of an interval from that marking its end. Pregnancy or birth histories typically are obtained through surveys rather than censuses, because they entail sequences of questions which are complex, variable from respondent to respondent, and potentially lengthy. Some data may also be available from vital registration systems where these, in the course of registering a birth, gather dates of previous live births and of marriage.

Analysis of birth intervals is most simply accomplished by calculating means or medians, the latter having the advantage of being less affected by a few very long intervals. It is, however, significantly complicated by ***three methodological problems***. The first is ***data quality***, which can be adversely affected by the misdating of births or failure to even remember their occurrence. The latter may in particular be a problem where children had died by the time of data collection, and especially where they had died very young. The second problem is ***censoring***, the reality that at a date of data collection there are always open birth intervals that may later become closed. Censored intervals that ultimately will be closed tend to be longer than already closed intervals of the same birth order, because the longer a birth interval is, the greater is the chance of it still being open at any given date of data collection. This means that for cohorts (age groups) of women yet to complete childbearing, mean or median birth intervals invariably are underestimates of their eventual values. In turn this obviously poses problems for comparison of cohorts that have, and have not completed their fertility. But even among the latter group the prospects of eventual closure of open intervals vary with age (a 20–24 year-old woman with an open third birth interval is more likely to close it than is a 45–49 year-old). One way of tackling this problem is to ignore more recent data for older cohorts so that cohorts effectively are compared at equivalent points in the life cycle. The other solution is to use life table techniques in which time elapsed since attaining parity n is the equivalent of age, and attaining parity $n + 1$ is the equivalent of death, in a conventional life table. This allows proportions attaining parity $n + 1$ within nominated birth intervals to be compared. The third methodological problem is ***selectivity of the birth intervals available for analysis*** in survey data. Because surveys typically target respondents in a restricted age range, intervals begun at any single year of age by persons outside the survey age range (often those no longer of reproductive age who have completed their fertility) are excluded from analysis.

In populations in which the date of first marriage is a good indicator of the commencement of exposure to the risk of conception, and in which practically no contraception is practiced until at least one live birth occurs, the first birth interval is useful for studying fecundability. Higher order intervals are less useful because couples with lower fecundability are selected out of them. Populations conforming to this model also tend to have shorter first than subsequent mean birth intervals, principally because after a birth the risk of pregnancy is reduced by post-partum amenorrhoea and the effect thereon of breastfeeding, customs prescribing post-partum abstinence (possibly), and circumstances such as maternal and infant health problems, reduced coital frequency, adoption of contraception and increased subfecundity.

Analysis of birth intervals often focuses on how they vary by characteristics of women, both social (e.g., education and labour force status) and biological or physiological (e.g., duration of breastfeeding and use of contraception). Obviously the more closely spaced children are the larger is the number who can be borne in a given period. The length of birth intervals thus affects the ***potential intensity*** of fertility, although this relationship is of limited practical moment in populations where even moderately high levels of fertility control are practiced. Research generally shows that the spacing of births has little impact on fertility ***levels***, the key determinant of which is the parity at which women decide to cease childbearing. Women can proceed quite rapidly to low parities then simply stop having children. Analysis of birth intervals by social characteristics thus may say more about the priorities of different groups of women through the life cycle than about their likely completed family sizes.

Birth intervals are also of interest in studies of child survival and maternal health. Very close spacing of births generally is found to be injurious to the health and survival chances of both mothers and their babies, and family planning programs often seek to promote wider spacing primarily for the health benefits likely to flow to women and children. The monitoring of family planning programs thus often entails monitoring their impact on birth intervals.

Whether dependent or independent, lengths of birth intervals have been prominent variables in demographic analysis in recent years. Much analysis has focused on specific intervals, and in particular the ***last closed birth interval*** (i.e., the interval between a woman's two most recent live births, or the first birth interval in the case of women with only one child). This interval yields the most up-to-date data on fertility and thus permits a focus on 'current' fertility.

Bogue (1993) has noted that, arithmetically, mean birth intervals are closely approximated by inverting marital fertility rates, and proceeds to ask what, in view of this, birth interval analysis has to offer. He concludes that it has three main virtues. First, it permits study of person-to-person variation in childbearing and of personal childbearing patterns, as distinct from the aggregate picture fertility rates provide. Second, because it is founded in survey research, it allows for the impact on childbearing of a much wider range of independent variables to be studied. And third, it facilitates subdivision of the total childbearing span into segments that are individually worthy of study – the interval from marriage to the first live birth;

intervals between successive live births; and the open interval from the last live birth to menopause.

Parity Progression

As has been previously noted, ***parity*** refers to the number of live births a woman has had. An alternative term sometimes used for the same concept is ***issue***. As well as being an attribute of women who have passed menarche, though, parity is also an attribute of any live birth, being the order of that birth (first, second, etc.) in the reproductive history of the woman who had it.

Parity is a particularly important variable to consider when analysing the fertility of populations that exercise significant control over their reproduction. Couples in such populations tend to have desired, or target, family sizes, sometimes a little flexible based on the gender mix of early births, but the attainment of which results in concerted efforts to avoid further births and thus, in the aggregate, in increasing departure of fertility from a 'natural' schedule as female age increases. A focus on parity allows the researcher to study the pattern produced by these efforts at fertility limitation, and changes therein as fertility preferences and/or the means of translating them into achieved family sizes change.

A number of the fertility measures discussed so far have been measures of some sort of 'average' experience. These, however, mask ***ranges*** of ***actual*** experience. A focus on parity highlights this variety of experience, and it not infrequently turns out that only a minority of women actually have approximately the 'average' number of children (to the nearest whole child). McDonald (1990) showed for Australia, for example, that if the 1988 pattern of fertility by birth order were to apply over the lifetimes of a cohort of women, only 24 % of those women would have two children (the 1988 TFR was 1.84), and only 26 % of children would live in two-child families. Thirty-two percent of women would have three or more children and 61 % of children would have at least two siblings. Contrary to what a TFR of 1.84 might conjure up in one's mind, while 24 % of women would have one child (20 % would remain childless), a mere 13 % of children would be only children.

Parity data are gathered by censuses, vital registers and surveys. Census data usually are obtained from a question like 'How many children have you ever had?' or 'How many children have you ever given birth to?'; i.e., from a question designed to yield the number of children ever born alive (CEB) to each woman. The question not infrequently is asked only of currently married or ever married women, since in many populations to ask young never married women whether they have had children is considered inappropriate. It often is accompanied by an instruction to exclude stillbirths, and is susceptible to understating parity unless also accompanied by supplementary questions asking about children still living and those who have died (there can be a tendency to overlook children born alive, but now deceased – especially those who died soon after birth). In the case of parity data gathered through vital registration it is wise to check the relevant question and instructions on the registration form. As previously noted Australian experience is that these may be worded with the aim of permitting ‘sensitive’ births (e.g., premarital births unknown to the father of the current child) to be ignored, perhaps by asking only about previous children ***of the current marriage or relationship***. The exclusion of some premarital births obviously introduces a bias here, but the consequences of the approach are even more serious in a population in which divorce and repartnering are common. If followed to the letter it results, for example, in a woman registering a first child by a second husband being classified parity one, when if she also had children by her first husband she is of higher parity. Parity data gathered in surveys may derive from census-type questioning or from pregnancy or birth histories.

While demographers have been known to calculate ***parity-specific*** (as distinct from order-specific) ***fertility rates*** (ratios of births of parity n during a year to the mean, or mid-year, female population aged 15–49 of parity $n-1$), the main application of parity data is to the calculation of ***parity progression ratios***. A parity progression ratio (PPR) is ***the probability that a woman who attains parity n goes on to attain parity n + 1***. Thus we refer to PPRs from parity zero to parity one, from parity one to parity two, etc. The use of the word ‘probability’ here is worthy of note. It is the first time it has occurred in this chapter, after having been frequently used in Chap. 4 in discussing the analysis of mortality. Recall our earlier distinction between renewable and non-renewable events, and that only the latter are suited to analysis using probabilistic measures. Childbirth is a renewable event, but in splitting births by parity we transform them into a series of ***non***-renewable events, thereby opening them up to probabilistic modes of analysis.

While it is possible to construct period, or synthetic cohort, PPRs (see Pressat (1972: 228 *et seq.*), they are more usually calculated for real female ***birth cohorts*** (i.e., age groups) or ***marriage cohorts*** (i.e., duration of marriage groups). In the case of birth cohorts PPRs may be calculated for all women, currently married women or ever married women, with data availability often dictating choice among these options. In general terms the PPR for a cohort from parity n to parity $n + 1$ is given by:

$$\mathrm{PPR}_{n,n+1} = a_n = \text{Women with at least } n+1 \text{ CEB/Women with at least } n \text{ CEB}$$

Given the distribution of a female birth or marriage cohort by number of CEB, the calculating equation for any PPR from parity n to parity $n + 1$ is:

$$a_n = \Sigma_{i=n+1,\omega}F_i / \Sigma_{i=n,\omega}F_i \tag{6.42}$$

Where $n =$ the parity ***from*** which women are progressing; $\omega =$ the highest parity attained by any member of the cohort; $F_i =$ the number of females of parity i in the cohort.

An example of the calculation of PPRs from a distribution of the ever married female birth cohort aged 45–49 by number of CEB at the 1990 Census of Thailand is presented in Table 6.8. Also shown are PPRs calculated in similar fashion for 45–

Table 6.8 Calculation of parity progression ratios for ever married female cohort aged 45–49 at 1990 census of Thailand, and comparison with ratios for Thai (1960) and ever married New Zealand (1981) cohorts

				a_n-values		Parity distribution per 1,000		
PPR calculations Thailand 1990				Thailand	NZ	Thailand	Thailand	NZ
Parity (n)	F_n	$\sum_{i=n,\omega} F_i$	a_n	1960	1981	1990	1960	1981
0	45,045	1,174,224	0.962	0.968	0.941	38	32	59
1	98,777	1,129,179	0.913	0.940	0.935	85	58	61
2	198,459	1,030,402	0.807	0.929	0.756	169	65	215
3	244,703	831,943	0.706	0.913	0.617	208	73	254
4	213,274	587,240	0.637	0.888	0.532	182	87	192
5	154,340	373,966	0.587	0.856	0.538	131	98	101
6	96,515	219,626	0.561	0.819	0.572	82	107	51
7	55,289	123,111	0.551	0.774	0.601	47	108	27
8	31,740	67,822	0.532	0.723	0.599	27	103	16
9	16,212	36,082	0.551	0.663	0.607	14	91	9
10+	19,870	19,870				17	178	15
Total	1,174,224					1,000	1,000	1,000

Sources: 1960 and 1990 Thai censuses; 1981 New Zealand census

49 year-old Thai females in 1960 and ever married New Zealand women of the same age in 1981, together with parity distributions for radix populations of 1,000 women in each of the three cohorts. The latter are presented as aids to comparison. They are obtained by first multiplying the radix l_0 by a_0 to determine the number of women who reach parity 1, l_1; then l_1 by a_1 to determine l_2; l_2 by a_2 to determine l_3; and so on. For each parity n we then take $l_n - l_{n+1}$ (for parity 10+ treat $l_{n+1} = 0$), the result being the number of women in the radix who finished childbearing at parity n. The procedure is the equivalent of finding d_x values in a standard life table.

The input data to the calculations for Thailand in 1990 are the column of F_n values (i.e., the census parity distribution for the cohort). In the next column these values are successively cumulated from the ***bottom*** of the column upward (i.e., F_{10+}, $F_{10+} + F_9$, $F_{10+} + F_9 + F_8$, etc.). Values of a_n (PPRs) then are found in accordance with Eq. 6.42 by dividing the cumulated value in each row of the table into the cumulated value in the row below it.

Several points concerning Table 6.8 deserve comment. First, the PPR calculations for Thai females in 1990 are based on women ***whose parities were known***. Parity was ***un***known for a further 50,392 ever married women aged 45–49. Ignoring these women effectively pro-rates them to the various parity categories, but conceivably they were disproportionately of zero parity. Assuming them to have ***all*** been of zero parity reduces a_0 from 0.962 to 0.922, and increases the number of parity zero women per 1,000 cohort members from 38 to 78 (with smaller compensating adjustments to numbers at higher parities). The issue of how ‘parity not stated’ cases are treated is, therefore, potentially a significant one, particularly if the degree of childlessness in a cohort is of interest. In the results shown for New Zealand in Table 6.8 the effect of ignoring such cases is less serious. Assuming all 'parity not stated' women to have been of zero parity reduces a_0 from 0.941 to 0.933, and increases the number of parity zero women per 1,000 cohort members from 59 to just 67.

Second, note that the calculations presented are for women aged 45–49; i.e., for birth cohorts ***that had completed their childbearing***, or were so close to doing so that they would have few subsequent births. Parity data for women still of reproductive age are censored as data on birth intervals are; further progressions to higher parities will take place in the future, and PPRs are biased by failing to take them into account (and not necessarily downward – if the earliest members of a cohort to attain a parity are the most likely to progress beyond it, the bias could be upward). Thus PPRs are of most use when calculated for cohorts that have completed their fertility, and if calculated for other cohorts must be interpreted with great caution. The situation when dealing with birth cohorts is reasonably clearcut – are cohort members or are they not beyond, or almost beyond, the age at which births can occur in significant numbers? With marriage cohorts, however, assessing whether reproductive potential of consequence remains is less straightforward, and needs to take account of ages at marriage. Except in exceptional circumstances, a cohort married 30 years or more will be beyond having many further children, but what about one married 15 years or 20 years? If tending on average to have married relatively late one could have more confidence in biology limiting further fertility than if having married early, but then ages at marriage follow a ***distribution***, and in almost any marriage cohort there would be ***some*** early marriers. Sometimes it is possible to examine parity progression for a marriage or birth cohort still with reproductive potential on the basis that that potential is likely to be realized mainly in higher order births, so that PPRs between the major, lower, parities are likely to be fairly reliable. But the issue of whether there is any censoring of cohort experience should always be carefully assessed.

Third, the 1960 and 1990 Thai schedules of PPRs (and parity distributions per 1,000 cohort members) in Table 6.8 are very different. The fact that the former pertains to all women and the latter to ever married women is of little consequence, since marriage among the cohort aged 45–49 in 1960 was well-nigh universal. Note that in the 1960 schedule PPRs decline ***gradually***, but ***consistently***, as parity increases. This tends to be the pattern in populations that exercise limited control over their fertility and in which biological processes (including deaths of husbands) play a major role in lowering fertility over the course of the reproductive life cycle. The parity distribution that results is heavily weighted towards the higher parities, with a large proportion of women having 10+ children. In the 1990 schedule, PPRs fall more rapidly to a level at about a_6 that is well below a_9 in the 1960 schedule, and at which they remain roughly constant thereafter. Clearly between 1960 and 1990 control over higher parity births increased considerably in Thailand, as is reflected in a parity distribution in which only 187 women per 1,000 had reached parity 6 or higher, compared to 587 per 1,000 for the equivalent cohort in 1960.

Fourth, the pattern of rapid decline in a_n with increasing parity to a level beyond which it becomes more or less constant is even more pronounced in the schedule of PPRs for ever married 45–49 year-old New Zealand women than in that for ever married Thai women of that age in 1990. Decline ceases at a_4. Such a pattern tends to be an empirical regularity in populations that practice high levels of fertility control, and this regularity has considerable potential for estimating ultimate completed fertility rates (cohort TFRs) for birth, and more especially marriage, cohorts that are only part way through their reproductive lives.

If a cohort has completed its childbearing there is a natural relationship between its PPRs and the cohort TFR. It is given by:

$$\text{Cohort TFR} = a_0 + a_0a_1 + a_0a_1a_2 + a_0a_1a_2a_3 + \cdots + a_0a_1a_2 \ldots a_{\omega-1} \qquad (6.43)$$

Where $\omega =$ the highest parity any cohort member attains.

A related equation allows us to estimate a cohort TFR for a marriage cohort married for a reasonable length of time (say 15 years or more) or a birth cohort the great majority of whose members similarly have been married for some time, but which is yet to complete its childbearing. It is only usable in respect of ***populations that exercise considerable control over fertility***, but increasingly many developing country populations are approaching this point. The equation is:

$$\text{Cohort TFR} = a_0 + a_0a_1 + a_0a_1a_2 + (a_0a_1a_2a_3/\,(1 - a_{4+})) \qquad (6.44)$$

Where $a_{4+} =$ the average of a_4, a_5 and a_6.

Table 6.9 shows data to which Eq. 6.44 might be applied. One hesitates to apply it to the cohort married for 10 years because the constancy of PPRs beyond a_3 is problematic, and equilibrium thus seems yet to be established. The basis of the method is an assumption that after a reasonable period of exposure to the risk of marital childbearing a cohort will have largely completed its progression through lower parities, with subsequent childbearing likely to be heavily concentrated at higher parities at a constant rate of parity progression. Ten years of exposure in an Australian setting in which couples routinely delay the first birth after marriage to establish married life on two incomes seems not to be enough to fully embrace this ‘progression’. But 15 years seems adequate, and applied to the cohorts married for 15, 20 and 25 years Eq. 6.44 yields estimated cohort TFRs of 2.69, 2.91 and 2.95 children per woman. This provides solid evidence of quite steep fertility decline from census data gathered around 15 years after that decline had begun.

PPRs shown in Tables 6.8 and 6.9 have been computed from distributions of female birth and marriage cohorts by number of CEB obtained from censuses. Similar distributions might as easily have come from surveys, and these are the types of sources you are likely to have to rely on in dealing with the populations of developing countries. However, data from reliable birth registers can also yield PPRs for birth and marriage cohorts. Without going into detail, we build up parity distributions for birth and marriage cohorts ***year by year*** using annual tabulations of live births by number of previous issue and age of mother (birth cohorts) or duration of marriage (marriage cohorts). There are, though, some complications to be aware

Table 6.9 Parity progression ratios for cohorts of Australian women married for 10, 15, 20 and 25 years, 1976

	Parity progression ratio							
Duration of marriage	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7
10 years	0.8991	0.8951	0.4722	0.2821	0.2358	0.2475	0.2848	0.3941
15 years	0.9203	0.9247	0.6318	0.4540	0.3598	0.3690	0.3587	0.3953
20 years	0.9105	0.9137	0.6945	0.5495	0.4526	0.4581	0.4625	0.4768
25 years	0.9117	0.8943	0.6880	0.5849	0.5092	0.5088	0.4925	0.5191

Source: Calculated from 1976 Australian census data

of. First, we must obviously obtain estimates of the ***original sizes*** of cohorts from somewhere to be able to estimate a_0, since a birth register gives no direct information on numbers of women who remain permanently childless. Possible sources are the birth register itself, a marriage register or a census, and it must be appreciated that computed cohort levels of childlessness will depend heavily on the accuracy of these estimates. Second, it may be necessary to make adjustments to allow for the changing size of cohorts over time due to migration, death and marriage dissolution. This can require a lot of additional data and be complicated. The beauty of basing PPRs on a census or survey distribution of a female cohort by number of CEB is that this sort of problem does not arise, because we are dealing only with 'survivors' at the date of data collection.

What have been covered in this section are the rudiments of parity progression analysis. With adequate data more complex descriptions of fertility processes are possible. One extension of the principles covered is to classify parity progression data by both age ***and*** duration of marriage, so that within birth cohorts the behaviour of different marriage cohorts can be contrasted and vice versa. It is also intuitively clear that there is a link between parity progression analysis and birth interval analysis. Together they fully describe the ***process*** of fertility – the former addressing the quantity or intensity dimension and the latter the timing or tempo dimension (relative to marriage as an initial reference point). Thus parity progression analysis can be combined with birth interval analysis. Indeed Feeney (1983) has shown how the two can be fused using ***parity progression schedules*** (consisting of the parity progression ratio and birth interval distribution for the transition from one parity to the next) and ***parity cohorts*** (groups of women attaining a given parity in a given period) into an approach to fertility analysis that (i) is more natural than that based on age-specific fertility rates (because it focuses on the ***whether*** and ***when*** to have a first or subsequent birth that is at the heart of fertility decision-making and because the proximate determinants of fertility operate by regulating parity progression and birth interval length), (ii) is able to overcome the technical problems of censoring and selection that affect simpler approaches to birth interval analysis, and (iii) provides a framework for looking at how birth intervals affect ***aggregate*** fertility and population growth trends, in contrast to micro-level analyses that concern themselves with what happens ***within*** birth intervals.

Biological Aspects of Fertility

From earlier discussion of the Bongaarts fertility model it will be clear that several of the proximate determinants of fertility are biological. In this section we deal briefly with those proximate determinants.

Fecundability

The term 'fecundability' refers to ***the probability of a fecund woman conceiving during a single menstrual cycle***. It is an important concept in statistical modelling of the family building process and in studying the effectiveness of family planning programs, even though Bongaarts sees natural fecundability as insufficiently variable from population to population to warrant incorporation in his model. Finding this probability might seem to require medically screening a panel of women to ensure they were fecund, regularly sexually active, and had no other impediments to their ability to conceive (e.g., sterility, short-term absence of their sexual partners, contraception, post-partum amenorrhoea, etc.). In many countries, however, ***newly married women who did not give birth before marriage and were not already pregnant at marriage*** largely meet these conditions, and one method of estimating fecundability focuses on this group. Be aware, though, that the method is inappropriate for populations that delay first births by practicing contraception immediately after marriage. It assumes sexual activity at an intensity unlikely to be exceeded subsequently in the marriage, lack of any attempt to prevent conception, and that both partners physiologically are never in future likely to be more fertile. Assessing fecundability in populations that practise contraception straight after marriage is much more difficult, and we will not attempt to deal with it here.

If the probability a newly married fecund woman will conceive during any menstrual cycle is f, the probabilities of conception during each cycle following marriage are given by the series:

$$f, (1-f)\ .\ f, (1-f)^2\ .\ f, (1-f)^3\ .\ f, (1-f)^4\ .\ f, \text{etc.}$$

That is, the probability of conceiving during the first cycle; the probability of ***not*** conceiving during the first multiplied by the probability of conceiving during the second; the probability of not conceiving during the first two cycles multiplied by the probability of conceiving during the third; and so on. After some mathematical manipulation it transpires that the mean number of cycles after marriage before conception, assuming that marriage coincides with ovulation, is 1/f, while assuming an even distribution of marriages through the menstrual cycle yields:

$$i = 1/f - {}^1\!/_2 \qquad (6.45)$$

Table 6.10 Calculation of fecundability of women who had given birth within 5 years of first marriage for respondents to the 1974 Malaysian fertility and family survey, by age at marriage

Age at marriage	Mean months to first birth	Mean months to first conception	Mean cycles to first conception	f
(1)	(2)	(3)	(4)	(5)
<15	26.0	16.5	17.9	0.054
15–17	20.1	10.6	11.5	0.083
18–19	17.5	8.0	8.7	0.109
20–21	16.5	7.0	7.6	0.123
22–24	16.5	7.0	7.6	0.123
25–29	18.1	8.6	9.3	0.102
30+	19.0	9.5	10.3	0.092

$(3) = (2) - 9.5$

$(4) = i = (3) . 365/336$

$(5) = 1 / ((4) + 0.5)$

Source: Malaysian fertility and family survey 1974, First country report, Table 2.1.1

Where i = mean number of cycles before conception; f = the probability of conceiving during a single cycle.

It follows that:

$$f = 1/\left(i + {}^{1}/_{2}\right) \tag{6.46}$$

Use of Eq. 6.46 to estimate fecundability for women who married at different ages is illustrated in Table 6.10. The input data are mean numbers of months between marriage and first birth for women who gave birth within five years of first marriage calculated from the Malaysian World Fertility Survey of 1974. Women with negative first birth intervals (implying premarital birth) and first birth intervals of less than 8 months (which imply premarital conception) are disregarded, the latter by adjusting published means on the assumption that births at marriage durations 0–7 months occurred on average at exact duration 4 months. Calculations then proceed through the following steps.

Step 1: Mean first birth intervals are changed to mean intervals to first conception by subtracting 9.5 months (column (3)).

Step 2: Mean intervals to first conception are converted to mean numbers of completed cycles before conception (column (4)) by multiplying by 365/336. This is approximately 13/12, and recognizes that a menstrual cycle lasts approximately four weeks, so that there are 13 in a year and each month to first conception embraces $1^{1}/_{12}$ cycles).

Step 3: Equation 6.46 is used to generate column (5) from column (4) (i.e., values of f from values of i).

Note that fecundability rises from a low level for women married when aged less than 15 to a peak for those married aged 20–24, then drops again for women married at older ages. This pattern seems intuitively plausible, allowing for ***adolescent subfecundity*** (a commonly observed phenomenon) and for declining fecundity at older reproductive ages. It is, however, necessary to remember that zero contraception prior to first conception is assumed. If, for example, women married at older ages were better educated, more likely to have careers, and consequently likely to have used contraception following marriage to delay first births, the drop in fecundability at those ages could, at least in part, be spurious.

Post-partum Amenorrhoea

A definition of amenorrhoea as the temporary absence of menstruation was given early in this chapter, where a distinction also was made between the condition during pregnancy (pregnancy amenorrhoea) and following confinement (post-partum amenorrhoea). The former is strictly biological and a constant in terms of its inhibiting effect on fertility, and thus is of limited interest to demographers. The latter, however, ***has a cultural dimension*** which makes its effect distinctly variable from woman to woman and from population to population. It is therefore of much greater interest both as a potential source of explanation for fertility trends and differentials, and as something family planning and maternal and infant health policy should take account of and possibly seek to influence. This cultural dimension has two components, one of which influences the biological condition itself and the other of which reinforces that condition, is capable of extending the period of post-partum non-susceptibility to the risk of conception, and is difficult to disentangle from the first. The former component focuses on ***cultural attitudes to breastfeeding***, which tends to extend the base period of post-partum non-susceptibility (estimated from studies of birth intervals following stillbirths to be about 3 months). Societies in which women typically breastfeed for extended periods exhibit longer intergenetic birth intervals, and hence lower completed family sizes under conditions of natural fertility. Breastfeeding is not a reliable means of pregnancy prevention at the individual level, but does seem to have some preventive effect for up to about 12 months following confinement so long as the child continues to receive breast milk as its only or main source of food (protection diminishes rapidly once other foods are introduced). Aside from the nutritional benefits to the child of breast milk, the longer birth intervals extended breastfeeding promotes are independently beneficial to the health and survival chances of both mother and child, and thus inherently worthy of encouragement through policy. They are also of value as a restraining factor on population growth and, insofar as they enhance child survival and confidence therein, as a factor in encouraging couples to stop childbearing at moderate parities.

The second component of the cultural dimension to the period of post-partum amenorrhoea is ***cultural restrictions on sexual intercourse during the post-partum***

period. Prescribed periods of abstinence from coitus may extend for fixed lengths of time, may be linked to the duration of breastfeeding through beliefs, for example, that semen will pollute the mother's milk, or both. Strictly speaking amenorrhoea is a biological condition, which breastfeeding extends through biological processes but which abstinence does not. Abstaining prevents conception by excluding semen from the female's vagina, not by preventing the resumption of ovulation. But abstention in conjunction with breastfeeding reinforces the biological condition of post-partum amenorrhoea, and if extending beyond the end of breastfeeding effectively prolongs the period of post-partum non-susceptibility to conception. Both together help determine when exposure to the risk of conception resumes; their individual effects on fertility are difficult to separate; and they thus are apt to be treated in tandem.

The measurement of post-partum infecundability, or the extent to which breastfeeding and post-partum abstinence combine with base-level post-partum amenorrhoea to defer the resumption of exposure to the risk of conception following a birth, was covered earlier in discussing Bongaarts's fertility model. Refer to Eqs. 6.39 and 6.40.

Sterility

Once again as already noted early in this chapter, sterility, or infecundity, are terms for a biological incapacity to conceive. The most straightforward measure of the extent of sterility in a population is the proportion of married women who reach menopause (approximated as ages 45–49) without having had a live birth. Such women are not necessarily physiologically sterile – they may be capable of conception but have miscarried due to other physiological problems – but this sort of measure is generally the only one readily available and is regarded as reasonable given that the alternative is an expensive special survey incorporating medical examinations. It is, of course, only of any use for populations where the desire for children within marriage is universal. If contraception is readily available and there is reason to believe some married women use it to remain voluntarily childless, then the measure becomes one of this voluntary childlessness and sterility combined (excluding, of course, voluntary childlessness associated with decisions not to marry), but is not a measure of sterility alone.

Sterility has a variety of types and causes, and is not necessarily a (reproductive) lifelong condition. Aside from lifelong sterility it can be caused by various diseases, including sexually transmitted diseases (STDs) such as gonorrhoea and chlamydia which sometimes lead to pelvic inflammatory disease (PID), and thence to permanent scarring of the fallopian tubes so that the passage of ova is prevented. Some strains of tuberculosis can have the same effect, and PID can arise from causes other than gonorrhoea and chlamydia. Other STDs, like syphilis, affect the viability of the foetus rather than fecundity, and thus increase the probability of miscarriage or stillbirth. Poor nutrition can induce ***temporary sterility*** by causing amenorrhoea or other menstrual irregularities, but restoration of appropriate nutrition usually quickly eliminates such a condition. Lifelong sterility is often referred to as ***primary sterility***, and that which prevents a woman having further births after she has already had one or more as ***secondary sterility***. It is, of course, only the former that the measure discussed in the previous paragraph is able to capture. Secondary sterility should also be differentiated from ***sub-fertility***, which refers to the situation of women who have extreme difficulty conceiving from the outset, although distinguishing between these conditions in practice is not easy and both may have similar connections to disease patterns in a society.

Finally, sterility can also be medically caused. Sometimes this occurs as a by-product of a medical procedure carried out for other than contraceptive reasons (e.g., a hysterectomy), but increasingly in recent decades couples and individuals have resorted to medically caused sterility as a terminal method of contraception. Either female (by tubal ligation) or male (by vasectomy) sterility may be sought to this end.

References

Bogue, D. J. (1993). Introduction to pregnancy/birth interval analysis. In D. J. Bogue, E. E. Arriarga, & D. L. Anderton (Eds.), *Readings in population research methodology* (Vol. 3, pp. 10-59–10-64). Chicago: Social Development Center for UNFPA.

Bongaarts, J., & Potter, R. G. (1983). *Fertility, biology, and behavior: An analysis of the proximate determinants*. New York: Academic.

Carmichael, G. A. (1986). Birth order in Australian vital statistics. *Journal of the Australian Population Association, 3*(1), 27–39.

Carmichael, G. A. (2013). Estimating paternity in Australia, 1976–2010. *Fathering, 11*(3), 256–279.

Coale, A. J., & Trussell, T. J. (1974). Model fertility schedules: Variations in the age structure of childbearing in human populations. *Population Index, 40*(2), 185–258.

Coale, A. J., & Trussell, T. J. (1975). Erratum to Coale and Trussell (1974). *Population Index, 41*(4), 572.

Corr, P., & Kippen, R. (2006). The case for parity and birth order statistics. *Australian and New Zealand Journal of Statistics, 48*(2), 171–200.

Davis, K., & Blake, J. (1956). Social structure and fertility: An analytic framework. *Economic Development and Cultural Change, 4*(3), 211–235.

Feeney, G. (1983). Population dynamics based on birth intervals and parity progression. *Population Studies, 37*(1), 75–89.

Henry, L. (1961). Some data on natural fertility. *Eugenics Quarterly, 8*, 81–91.

Henry, L. (1976). *Population analysis and models*. London: Edward Arnold.

Jones, E. F. (1971). Fertility decline in Australia and New Zealand 1861–1936. *Population Index, 37*(4), 301–338.

McDonald, P. (1990). The 1980s: Social and economic change affecting families. *Family Matters, 26*, 13–18.

Newell, C. (1988). *Methods and models in demography*. London: Belhaven Press.

Pressat, R. (1972). *Demographic analysis: Methods, results, applications*. London: Edward Arnold.

Shryock, H. S., Siegel, J. S., & Associates. (1973). *The methods and materials of demography*. 2nd Printing (revised). Washington, DC: U.S. Bureau of the Census.

United Nations. (1983). *Manual X. Indirect techniques for demographic estimation*. ST/ESA/SER.A/81. New York: United Nations. Available online at http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html. Accessed 26 July 2013.

World Health Organization. (1993). *International statistical classification of diseases and related health problems* (Tenth Revision, Vol. 2). Geneva: WHO.

Chapter 7
Population Distribution, Urbanization and Migration

Demography and Geography

Certain themes in demographic analysis have a distinctly geographic dimension to them, and geographers not surprisingly have taken particular interest in them. This is not the place for a review of the development and contemporary nature of population geography (Jones 1990; Newbold 2010). Nor does one wish to create the impression that themes addressed in this chapter represent anything like the totality of the demographic issues and phenomena in which geographers take an interest. But the geographic (or ‘spatial’) distribution of population, the movement of population between geographic locations (variously mobility and migration), and the process of urbanization, to which population redistribution by migration invariably is fundamental, are all inherently geographic, and hence clearly of central concern to geographers.

Population Distribution

Thus far in this book populations focused on in illustrating demographic techniques and methodological principles have been ***national*** populations, or subgroups thereof defined by variables like ethnicity. For many purposes, however, national-level data are too crude. The geographic distribution of national populations typically is very uneven, with large numbers of people living in relatively small areas in urban centres and sometimes, at the other extreme, vast areas containing almost nobody. Australia is a classic example, with its population heavily concentrated in cities and towns along the eastern, southeastern and southwestern coastal areas, and huge inland tracts that are virtually uninhabited. Then again, demographic characteristics and processes rarely occur uniformly over a country; subnational

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_7

political, administrative and planning bodies have responsibility for, and therefore an interest in, restricted geographic areas; and these same bodies and national governments are concerned to know about population movement ***within*** as well as across national borders. All of these considerations argue for demographic data tabulated for discrete geographic areas within a country. Such data normally derive from censuses, and to a lesser extent from vital registers, population registers and administrative sources. They are less likely to come from sample surveys, although national-level sample surveys can sometimes yield broad regional subsamples and frequently allow urban/rural and/or metropolitan/non-metropolitan distinctions to be made.

Levels at Which Data Are Tabulated

Most countries of any size tabulate demographic, and especially census, data at a range of different geographic levels. Statistical agencies usually must recognize as a priority in this process a country's ***political or administrative subdivisions***, which may range from primary units such as the states of Australia, the USA or India and the provinces of Canada, Indonesia or Thailand through a hierarchy of subsidiary levels to local administrative units which may exist at the village or sub-village level in rural areas, and at the suburban or neighbourhood level within cities. In many countries simply counting numbers of persons, or of persons with particular attributes, resident within these subdivisions is ***the*** major reason for conducting a census at all, such counts determining the geographic allocation of certain resources and apportionment of parliamentary representation. Population counts for political or administrative units also are necessary because those units are natural foci for planning, social and economic policy development, and studies of internal migration.

Hierarchies of administrative units vary in the number of levels they have and in the names given to units at each level, and often bifurcate along urban-rural lines. Thus cities, towns, villages and various administrative subdivisions thereof (boroughs, wards, *arondissements* (France), *ku* (Japan, Korea), etc.) often are recognized alongside a system of counties, shires, etc. which cover rural areas (and may overlap with the urban system at lower levels). Statistical agencies may also, however, create geographical units ***for purely statistical purposes***. Primary administrative units sometimes are aggregated into broader regions for statistical purposes, as occurs, for example, with the publication by Indonesia of data for island groupings (which are aggregates of provinces). Regions may also be designed which are perceived to be ***functional economic or cultural areas***. These may be groupings of secondary or tertiary administrative units, may cut across the boundaries of primary administrative units, or may bear no relation to administrative boundaries at all. Census authorities also are often keen to draw and periodically revise their own boundaries around urban agglomerations so as to reflect functional rather than political or administrative reality, and may create areas which facilitate data collection, cater to the demand for small-area data (see below), are relatively homogeneous (particularly socio-economically or culturally), impart flexibility to the system of geographical classification, and facilitate the presentation of data at levels of geographic disaggregation perceived to be useful across both urban and rural landscapes. Thus in Australia, for example, under the new Australian Statistical Geography Standard (ASGS) introduced in 2011, the basic spatial building block for release of census data is the Mesh Block, for which little more than population and dwelling counts are produced. Around 350,000 of these aggregate to 55,000 Level 1 Statistical Areas (SA1s) with populations in the range 200–800 people (the smallest areas for which more extensive census data are released); these aggregate to 2,200 Level 2 Statistical Areas (SA2s) with populations between 3,000 and 25,000; these in turn aggregate to around 350 Level 3 Statistical Areas (SA3s) of 30,000–130,000 people; then there are just over 100 Level 4 Statistical Areas (SA4s) with populations between 100,000 and 500,000 which aggregate to the nine States and Territories and then to Australia. Various levels in this hierarchy also aggregate to a range of other regional subdivisions, including such things as Postal Areas, State and Commonwealth Electoral Divisions, Tourism Regions, Significant Urban Areas, and Greater Capital City Statistical Areas.

Mention was just made of ***small-area data***. Particularly for urban areas, such data tend to be in heavy demand, and the smaller the small areas the better so far as many users are concerned. In some countries the lack of any administrative unit sufficiently small to meet small-area data requirements has been a major reason for developing geographic units which are first and foremost statistical entities, and only incidentally ever administrative ones. Small-area data permit one to focus on localized communities and to differentiate urban socio-demographic landscapes in considerable detail. They also offer great flexibility in constructing regions to personal specification through aggregation. All of these features are very attractive to planners, businesses making location decisions and marketers. It should be appreciated, though, that as geographic disaggregation increases, the classificatory and cross-tabulatory detail in data statistical agencies are prepared to release may tend to diminish, as the confidentiality of individual respondents becomes more of an issue.

Urban and Rural Populations

Significant differences long have been recognized to exist between populations that live in cities and in small villages or the open countryside. Hence differentiating between urban and rural populations in demographic data collections has a lengthy history. But if there is broad agreement on the ***desirability*** of making an urban-rural distinction, there is no universally accepted ***means*** of making it. The extremes are not a problem. The hearts of major cities are clearly urban and areas in which people almost universally reside on and work on farms are rural. But where, along the continuum in between, does urban stop and rural begin?

The upshot is that criteria used to separate urban from rural populations vary considerably between countries, and may vary over time within them. There is not even agreement that a standard definition of ‘urban’ is ***desirable***, it being widely felt that different types of definitions suit different purposes and types of countries. Thus, when dealing with urban-rural data routinely acquaint yourself with the definition of ‘urban’ that underpins them, especially if any comparative analysis is to be undertaken (note that ‘rural’ generally is not directly defined, but is treated as a residual after eliminating ‘urban’).

The recommended focus when dividing populations into urban and rural components is the ***locality***, defined by the Statistics Division of the United Nations (SDUN) (2008: 123) as ‘a distinct population cluster in which the inhabitants live in neighbouring sets of living quarters and that has a name or a locally recognized status.’ If that is not possible, the smallest administrative unit in a country should be used. Each locality or administrative unit, and the population residing at or within it, is determined to be urban or not urban, the populations of non-urban localities and areas not deemed to be localities being the rural population. After affirming that cross-national differences in the characteristics that distinguish urban from rural areas make a universal definition of ‘urban’ impractical, the SDUN goes on to note that, in industrialized countries, lifestyle and level-of-living lines of differentiation between urban and rural areas have become blurred, leaving the degree of concentration of population as the main discriminator. Thus ***population size and density*** generally are key considerations in classifying localities as ‘urban’. There is, however, no consensus as to appropriate minimum values for these variables, which tend to vary depending on how populous and densely settled a country is overall. In less developed countries, where large densely settled areas still are characterized by ‘a truly rural way of life’, additional criteria, and perhaps something more elaborate than a simple urban-rural classification, are in order. Among these extra criteria are the ***extent of employment in non-agricultural activities*** and the ***presence of urban facilities***. The latter include such things as access to electricity, piped water and a flush toilet in the home, and to health services, education and a local public transport system.

Other criteria by which localities may be classified ‘urban’ include ***whether or not they have been classified ‘urban’ administratively*** by government and ***whether they are subject to some form of local government***. Classifications in which these principles feature have a tendency to be unhelpfully static and outdated. Thus, for example, significant areas of clearly urban character by objective criteria may be excluded from the urban population because the wheels of bureaucracy turn slowly in formally designating them urban or adding them to existing ‘urban’ areas with which they are contiguous. Thailand is a present day case in point, with scholars routinely having to adjust official estimates of the urban population upward to achieve a realistic outcome.

Methods of Analysis

A seemingly very straightforward approach to examining population distribution is to ***map the numbers or densities of inhabitants*** of geographic areas or localities. A common method of mapping population numbers is to use a dot to represent a fixed number of people (say 100 or 1,000), dots being located physically within areas to which they pertain. This method can, however, run into difficulties when both sparse rural and dense urban concentrations of population need to be portrayed. Dots representing urban populations can merge into an undifferentiated mass. One solution is to use dots of a different colour and representing a larger number of people in urban areas. Another is to represent urban localities by a geometric image (perhaps a circle) of size proportional to population.

If mapping population densities the usual approach is to develop a set of density categories, associate with each a distinctive shading in a scheme in which density of shading increases with density of population, and shade each geographic area according to the category into which its population density falls. Well designed and executed, maps of this type can have a striking visual impact, although they do sacrifice some precision. Population densities are calculated by dividing the population of an area by its physical area in, for example, square kilometres to yield a figure of so many persons per square kilometre. They can, however, be deceptive. If large parts of a country or region are physically inhospitable and therefore largely uninhabited, densities may bear little relation to actual densities in those areas that are populated to any significant extent. In this circumstance density calculations may for some purposes usefully be made with reference to ***arable agricultural*** land area rather than to ***total*** land area. Urban population densities can also deceive, depending on the definition of 'urban' employed. The more a city is defined to include urban-rural fringe areas the lower its population density is likely to be.

Another simple way of summarizing population distribution is to list the geographic units of a country or region and calculate ***percentages of total population living in each***. Such tabulations facilitate the measurement of change in the distribution of population through comparison of percentage distributions at discrete time points. Normally we focus on a level of geographic disaggregation with relatively few categories (e.g., the States and Territories of Australia), so that individual percentages, and changes therein, do not become so small as to mask major trends. Geographic localities may also be ***rank-ordered by population size***, a strategy again lending itself to examining distributional change over time. Comparison of rank orderings at two censuses can point to areas that have grown comparatively rapidly or slowly in the interim, although this procedure may be insensitive at higher ranks in the population size hierarchy, where rank to rank differences in population are so large that substantially different intercensal growth rates would produce no change in ranking.

Statisticians and geographers have from time to time proposed measures of the ***average location*** and ***degree of dispersion*** of population distributions, although they are somewhat peripheral to the technical armoury of the specialist demographer.

Locating the ***median point*** of a population distribution involves drawing the east-west line on a map above and below which exactly half the population is located, and the north-south line that also divides the geographic distribution in two. The median point is then the intersection of these two lines. Intersections of east-west and north-south lines that divide distributions into quartiles, deciles etc. can also be located. Plotting these points for successive census dates can provide an interesting perspective on distributional change over time.

A variant on the median point of a population's geographic distribution is the ***centre of population***, otherwise known as the ***mean point*** of the distribution or its ***centre of population gravity***. After overlaying an x,y-scaled grid on a map of population distribution, the (x,y) coordinates of the centre of population are given by:

$$x = \Sigma_{i=1,\omega} p_i \,.\, x_i / \Sigma_{i=1,\omega} p_i; \; y = \Sigma_{i=1,\omega} p_i \,.\, y_i / \Sigma_{i=1,\omega} p_i \tag{7.1}$$

Where p_i = the population of unit of area i; x_i and y_i = the x and y coordinates of the centre of area i; ω = the number of units of area into which the country or region being studied is divided.

Unlike the median point, the centre of population is influenced by the ***distance*** of each person from it. Population redistribution ***within*** a quadrant formed by intersecting east-west and north-south median lines has no effect on the median point, but does alter the centre of population. Again, movement of the centre of population over time captures change in the distribution of population.

One measure of the dispersion of a population distribution is the ***standard distance***. This bears the same relationship to the centre of population as the standard deviation bears to the arithmetic mean in statistics. While in theory it should be calculated based on individual members of the population and their locations relative to the centre of population, in practice we normally work, as in Eq. 7.1, with data giving population counts for (preferably quite small) units of area. Under this scenario the standard distance is given by:

$$D = \sqrt{}\left\{\left[\Sigma_{i=1,\omega} f_i (x_i - x)^2 + \Sigma_{i=1,\omega} f_i (y_i - y)^2\right] / n\right\} \tag{7.2}$$

Where f_i = the population of unit of area i; x_i and y_i = the x and y coordinates of the centre of area i; x and y = the x and y coordinates of the centre of population; n = the total population of the country or region for which D is sought; ω = the number of units of area into which that country or region is divided.

The answer obtained will be in units of distance corresponding to a value of 1 on the scales of x and y coordinates. It can be converted to a meaningful distance (e.g., in kilometres) by multiplying by whatever this unit of distance is (e.g., if an increment of 1 on the x or y scale corresponds to 100 km, we multiply by 100 to obtain an answer in kilometres).

A graphical technique for depicting the concentration or dispersion of population is provided by ***the Lorenz curve***. Originally developed for studying inequalities in the distribution of wealth or income it requires geographic localities within a country or region, or groups of localities classified by size of population, to be arrayed from the most to the least populous, and cumulative proportions of localities and of total population to be calculated down this array. These cumulative proportions then are plotted against one another (population on the x-axis; localities on the y-axis) and compared to the x = y diagonal, which represents the situation where all localities have equal population (maximum dispersion of population). The closer to this diagonal a Lorenz curve lies the less concentrated (or more evenly dispersed among localities) a population is. Note that where a proportion of population is reported as not living in localities it has to be ignored. This is not a problem in Fig. 7.1, which shows Lorenz curves for Venezuela and Italy in 1981; only 0.5 % of Venezuela's population and none of Italy's was not living in a locality. It can be seen that the Venezuelan curve hugs the x-axis and righthand border of the graph much more closely than that for Italy, or in other words is decidedly further displaced from the reference diagonal. It signifies a population a very high proportion of which is concentrated in very few localities, and in which a relatively high proportion of localities are home to a relatively small proportion of population. The pattern for

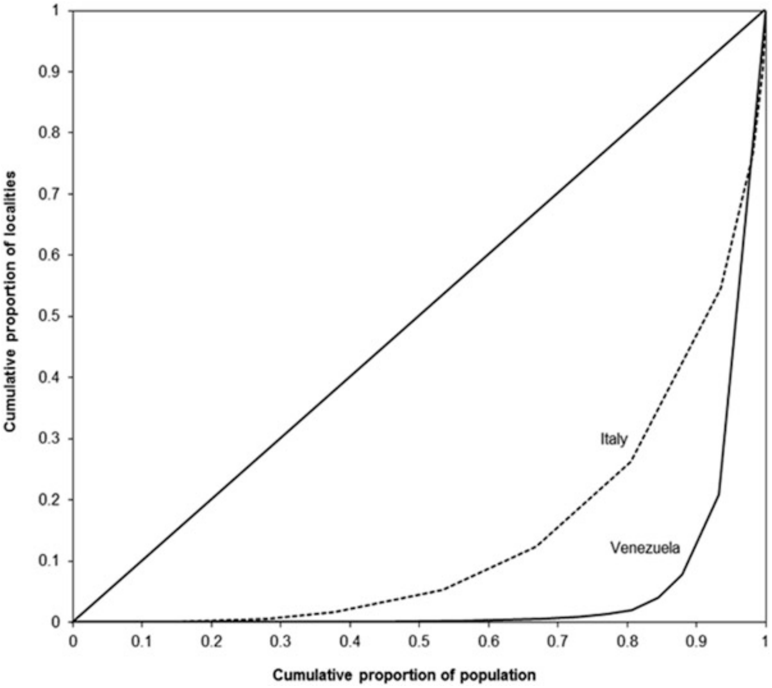


Fig. 7.1 Lorenz curves showing degrees of population concentration in Venezuela and Italy, 1981

Italy, while also indicating a concentration of population in relatively few, larger localities, is much less pronounced. Clearly there is a wider geographic spread of population.

The degree of concentration of population indicated by a Lorenz curve can be quantified. The ***Gini concentration ratio*** measures the proportion of the area under the reference diagonal that lies between the diagonal and the Lorenz curve. The higher the ratio the greater is the concentration of population in a few localities. The relevant equation is:

$$\mathrm{Gi} = \Sigma_{i=1,n-1} x_i . y_{i+1} - \Sigma_{i=1,n-1} x_{i+1} . y_i \tag{7.3}$$

Where x_i = cumulative proportion of population living in localities in size category i and all larger categories; y_i = cumulative proportion of localities in size category i and all larger categories; n = number of size of locality categories arrayed from largest (n = 1) to smallest (n = n).

Table 7.1 sets out the calculations for obtaining both the (x_i, y_i) Lorenz curve plots and the Gini concentration ratio for Venezuela. The Gini ratio of 0.8919 compares with a value of 0.7115 for Italy.

Another approach to analysing population concentration using the Lorenz curve and Gini concentration ratio entails focusing on ***areas*** rather than localities. Instead of arraying localities and their populations from the most to the least populous, a country's (or region's) ***geographic subareas***, together with their land areas and populations, are arrayed ***in order of population density*** from most dense to least dense. Cumulative proportions ***of total land area*** and of total population then are calculated down this array and plotted against one another as before (population on the x-axis; land area on the y-axis), and the Gini concentration ratio is calculated from these sets of (x_i, y_i) values as before. One significant advantage of this approach is that the entire population is covered. There is no problem with the sometimes sizeable population that is deemed not to be living in a locality and therefore has to be disregarded; all people reside in one or another geographic subdivision.

We will also deal here with a measure commonly known as the ***index of dissimilarity***. This is not a measure specifically developed to aid the study of population distribution, but it can be used for that purpose. The index of dissimilarity enables two percentage distributions to be compared, and indicates ***the percentage of either of the comparator populations who would have to change categories (be redistributed) for the two distributions to become identical***. Thus its value ranges from zero, where two distributions are already identical, to 100, where all members of one population are in different categories from all members of the second. The word 'population' is used here in the statistical sense – the percentage distributions do not have to be distributions of ***people*** (although in demographic applications they mostly will be). They do, however, have to feature ***the same categories*** in any comparison. You cannot compare a percentage distribution by age groups with one by, say, categories of religion (not that it would make any sense to try to do this),

Table 7.1 Calculation of Lorenz curve coordinates and Gini concentration ratio for Venezuela, 1981

					Cumulative proportion of			
			Proportion of		Localities	Pop'n		
Size of locality	No. of localities	Population	Localities	Pop'n	y_i	x_i	$x_i \cdot y_{i+1}$	$x_{i+1} \cdot y_i$
500,000+	3	2,559,517	0.0002	0.1772	0.0002	0.1772	0.0002	0.0001
100,000–499,999	19	3,688,847	0.0010	0.2554	0.0011	0.4326	0.0011	0.0006
50,000–99,999	28	1,905,009	0.0015	0.1319	0.0026	0.5645	0.0030	0.0018
20,000–49,999	53	1,610,324	0.0028	0.1115	0.0054	0.6760	0.0057	0.0039
10,000–19,999	57	776,551	0.0030	0.0538	0.0084	0.7298	0.0095	0.0065
5,000–9,999	89	598,963	0.0046	0.0415	0.0130	0.7713	0.0158	0.0105
2,000–4,999	143	496,010	0.0075	0.0343	0.0205	0.8056	0.0324	0.0173
1,000–1,999	377	569,642	0.0197	0.0394	0.0402	0.8450	0.0658	0.0354
500–999	722	496,523	0.0377	0.0344	0.0779	0.8794	0.1830	0.0726
200–499	2,493	758,630	0.1302	0.0525	0.2081	0.9320	0.9320	0.2081
<200	15,165	982,782	0.7919	0.0680	1.0000	1.0000		
Total	19,149	14,442,798	1.0001	0.9999			1.2486	0.3567
Gini concentration ratio $= 1.2486 - 0.3567 = 0.8919$								

and if comparing two age distributions the age groups would need to be the same in both distributions.

The index of dissimilarity is given by:

$$\mathrm{ID} = {}^{1}\!/_{2}\Sigma_{i=1,k}\,|x_i - y_i| \tag{7.4}$$

Where $k =$ the number of categories in each of the percentage distributions being compared; $x_i =$ the percentage in category i in the first distribution; $y_i =$ the percentage in category i in the second distribution.

If the notation which sees the expression $x_i - y_i$ enclosed by two vertical lines in Eq. 7.4 is unfamiliar, it means ‘the absolute value of’. This in turn means that if the expression enclosed turns out to be a negative number, ***the minus sign is disregarded*** and the number is treated as though it was positive.

One application of Eq. 7.4 to the study of population distribution is to comparisons over time of percentage distributions of population among geographic areas. The index of dissimilarity allows the degree of change between two dates to be quantified on a 0–100 scale (although realistically values are likely to fall towards the lower end of this range). The index can, however, also be used to measure population concentration. To do this a country’s, or a region’s, geographic subareas are the categories, and percentage distributions among them of ***total land area*** and ***total population*** are compared. Used in this way the index of dissimilarity is known as the ***index of concentration***. A value of zero indicates an absolutely even distribution of population to geographic subareas in proportion to their proportions of total land area (zero concentration). Higher values indicate the degree to which population is distributed at above average densities to some areas, and below average densities to others. In this application the maximum possible value is uncertain, but is less than 100, since in the limiting case all population would reside in one subarea. That subarea, however, could not have zero land area, and thus zero percent of total land area, so that not all the 100 % of population would need redistributing for the population and land area distributions to match.

Urbanization

The topics of population distribution and urbanization are closely interlinked. While definitions of urbanization vary, just as definitions of ‘urban’ do, there is little dispute that both the process and the level of urbanization in a population have broadly to do with geographic concentration of population. The level of urbanization can be defined as ***the percentage of total population living in areas defined to be ‘urban’***. It is thus susceptible to the vagaries of different definitions of ‘urban’, and these should always be assessed, especially when comparisons are being made. If population size is a key criterion, are urbanization levels for different populations based on the same minimum size threshold, and are the criteria used to define the boundaries of urban localities comparable? If the definition emphasizes administrative units defined to be urban, how up to date is the process of conferring urban status on areas newly deserving of it? To what extent do definitions for different populations place different emphases on these criteria and on criteria such as the prevalence of non-agricultural employment and access to ‘urban’ facilities?

Under a definition of urban that emphasizes the sizes of localities’ populations, the ***process*** of urbanization occurs due to (i) additions to the populations of urban places at rates in excess of additions to the populations of non-urban places, and (ii) formerly non-urban places attaining the minimum population threshold to be deemed urban (at a rate faster than formerly urban places slip below that threshold). These sorts of changes often occur mainly as a result of internal or international migration, cities being magnets for migrants. But natural increase can also be a factor, especially in carrying localities past a minimum population threshold. The physical growth of cities due to migration and natural increase also increases urbanization by capturing formerly non-urban areas and their populations. Others, however, see the process of urbanization being more complex, previously non-urban areas meriting reclassification not just on the basis of population size but on the basis of a change from largely agricultural to largely non-agricultural employment, or through the acquisition of ‘urban’ facilities.

Methods of Analysis

The most commonly encountered measure of the ***level of urbanization*** is that indicated above as defining that level – the percentage of total population living in urban areas. Thus:

$$pu = (u/t)\ .\ 100 \tag{7.5}$$

Where u = urban population; t = total population.

Where international comparisons are to be made the percentage of total population living in ***localities of at least some minimum population size*** (e.g., 20,000) is perhaps a preferable measure. This really amounts to imposing a uniform definition of ‘urban’, at least insofar as size of locality is concerned (the criteria used to establish locality boundaries are another matter), after which Eq. 7.5 can again be used. Other measures of the level of urbanization have been proposed which go beyond the simplicity of these two and try to take the sizes or size classes of urban localities into account (the extent to which urban population is concentrated in localities of larger size), but none has been widely adopted.

Besides the level of urbanization, analysts often measure the ***tempo of urbanization***. One common approach has been to use the compound interest formula to calculate the ***average annual rate of urban population growth*** between two points in time at which the size of the urban population is known. The compound interest formula is:

$$Q_2 = Q_1(1 + r)^n$$

Where Q_1 and Q_2 are quantities at times 1 and 2, n years apart; $r =$ the annual rate of growth in Q.

If Q_1 and Q_2 are urban populations n years apart which we denote as u_1 and u_2, this formula can be modified and rearranged to yield the average annual rate of urban population growth:

$$r_u = {}^n\sqrt{(u_2/u_1)} - 1 \tag{7.6}$$

Although widely used, the average annual rate of urban population growth has, for technical purists, drawbacks which disappear if an ***instantaneous rate of urban population growth*** is calculated. The equation for this growth rate derives from the exponential formula:

$$Q_2 = Q_1 \cdot e^{in}$$

Where Q_1 and Q_2 are again quantities at times 1 and 2, n years apart; $i =$ the instantaneous rate of growth in Q.

If Q_1 and Q_2 are urban populations n years apart which we again denote as u_1 and u_2, the exponential formula can be modified and rearranged to yield the instantaneous rate of urban population growth:

$$i_u = \ln(u_2/u_1)/n \tag{7.7}$$

In this equation 'ln' means 'the natural logarithm of'. You may be wondering if there is a link between r_u and i_u. There is, and in the notation of the compound interest and exponential formulae it is provided by the relationship $r = e^i - 1$. Using this relationship with $r = r_u$ and $i = i_u$, an alternative equation for r_u can be obtained:

$$r_u = e^{\ln(u_2/u_1)/n} - 1 \tag{7.8}$$

If urban population is increasing, the average annual growth rate r_u will be slightly higher than the instantaneous growth rate i_u, the difference increasing as the growth rates increase. If urban population is declining, the negative instantaneous growth rate i_u will be slightly larger (numerically) than the negative average annual growth rate r_u. Either growth rate will, however, serve the purposes of distinguishing periods of more rapid and less rapid urban growth for a population, and differentiating populations with more rapid and less rapid urban growth over a specified period of time.

Average annual and instantaneous rates of urban population growth are ***not***, however, strictly measures of change ***in the level of urbanization***. Urban population growth need not mean increasing urbanization if rural population is growing as, or more, rapidly, and even if urbanization ***is*** increasing it is unlikely to be doing so at the same rate as urban population. Thus, four other measures of the tempo of urbanization are the ***urban-rural growth difference***, the ***average annual increment in the level of urbanization***, and the ***average annual*** and ***instantaneous rates of increase in the level of urbanization***.

Average annual and instantaneous rates of ***rural*** population growth, r_r and i_r, can be found using Eq. 7.6 (or Eq. 7.8) and Eq. 7.7, respectively, by substituting the rural populations at times 1 and 2, r_1 and r_2, for u_1 and u_2. The urban-rural growth difference then is given by:

$$\mathrm{URGD} = r_u - r_r \tag{7.9}$$

Or:

$$\mathrm{URGD} = i_u - i_r \tag{7.10}$$

The average annual increment in the level of urbanization is given by:

$$i_u = (p_{u,2} - p_{u,1})\,/n \tag{7.11}$$

Where $p_{u,1}$ and $p_{u,2}$ = the percentages of population that were urban at times 1 and 2; n = the interval in years between times 1 and 2.

The average annual and instantaneous rates of increase in the level of urbanization, ri_u and ii_u, can be found using Eq. 7.6 (or Eq. 7.8) and Eq. 7.7, respectively, but writing $p_{u,1}$ and $p_{u,2}$ (with the same meanings as in Eq. 7.11) for u_1 and u_2 respectively. Thus:

$$ri_u = {}^{n}\!\surd\,(p_{u,2}/p_{u,1}) - 1 \tag{7.12}$$

Or:

$$ri_u = e^{\ln(p_{u,2}/p_{u,1})/n} - 1 \tag{7.13}$$

And:

$$ii_u = \ln(p_{u,2}/p_{u,1})\,/n \tag{7.14}$$

It is worth considering briefly each of these measures of the tempo of urbanization in the context of the ***projection*** of urbanization levels. The rate of urban population growth is normally higher than the rate of rural population growth, and therefore also higher than the rate of total population growth. It thus follows that, projected sufficiently far into the future, constant rates of urban and total population growth where the former exceeds the latter will eventually result in the absurdity of an urban population that exceeds total population. Similarly, projecting an annual increment in the level of urbanization into the future eventually produces the absurdity of a level of urbanization in excess of 100 %, while projecting an annual ***rate*** of increment in the level of urbanization leads to this absurdity even faster, because the annually increasing base level to which the rate is applied means an annually ***increasing***, as opposed to constant, increment in the level of urbanization. The one measure which does not inherently lead ultimately to an absurd result is the urban-rural growth difference, although that is not to say that projections based on it will be accurate.

Similar levels of urbanization can mask different ***urban structures***. Students of urbanization are also concerned to measure this dimension of a country's rural-urban distribution of population. Is urban population concentrated in a single large, or ***primate***, city, or does it tend to be distributed in a more balanced way among several large urban agglomerations and a network of smaller cities and towns? The former pattern tends to be found more often, but not always, in less developed regions; the latter in more developed countries.

One approach to studying urban structure, or the ***urban hierarchy***, is to examine absolute and percentage distributions of urban localities and urban population ***by locality size classes***. Such distributions may be studied cross-nationally or over time for individual countries to trace differentials in, or the evolution of, urban hierarchies. The Lorenz curve and Gini concentration ratio outlined in the previous section are also devices that permit us to focus on a country's urban structure. The Venezuelan example presented in Fig. 7.1 is typical of a strongly primate urban hierarchy, while the Italian example is illustrative of a much more balanced hierarchy.

Primacy may be measured by the ***percentage of urban population living in the largest city***. This measure is sometimes also treated in conjunction with a ***primacy index*** or ***four-city index***, which is the ratio of the population of the largest city to the sum of the populations of the second to fourth largest cities. Thus:

$$\mathrm{PI} = \mathrm{p}_1 / \left(\mathrm{p}_2 + \mathrm{p}_3 + \mathrm{p}_4\right) \tag{7.15}$$

Where $\mathrm{p}_1, \ldots, \mathrm{p}_4$ are the populations of the first through fourth largest cities.

A condition of primacy is said to exist if PI exceeds 1.0. Measures of primacy should always, however, be analysed in conjunction with measures of the level of urbanization and city size distributions, since very high primacy can coexist with quite low levels of urbanization (e.g., Conakry in Guinea was in 1990 home to 76 % of the nation's urban population, but only 26 % of the population was urban), and low primacy does not exclude the existence of very large cities (e.g., China in 1990 had a primacy index of only 0.543, but had two cities with populations over 10 million).

Urban structure can also be examined by dividing the urban population into 'city' and 'town' components based on size of locality, then treating these two components of urban population in ways analogous to those used to treat the urban and rural components of total population. Thus, the percentage of urban population living in 'cities' becomes a measure of ***the level of concentration of urban population*** analogous to the percentage of total population that is urban (the level of urbanization). Likewise measures of the tempo of urban concentration can be constructed which are analogous to the measures of the tempo of urbanization discussed above; the ***city-town growth difference***, the ***average annual increment in the level of concentration of urban population*** and the ***average annual*** or ***instantaneous rate of increase in the level of concentration of urban population***.

A benchmark sometimes used for the analysis of urban structures is the ***rank-size rule***. Although observed empirically by Felix Auerbach as early as 1913, it was fully developed by George Kingsley Zipf in the early 1940s. It states that the expected size of a city's population is inversely proportional to its rank in the urban hierarchy. That is:

$$p_i = k/r_i \tag{7.16}$$

Where p_i = the population of urban locality i; r_i = the rank of urban locality i in the urban hierarchy; k = the population of the largest city in the urban hierarchy.

Thus the second-ranked city in an urban hierarchy is expected to have half the population of the first-ranked, the third-ranked to have one-third of the first-ranked city's population, and so on. The rank-size rule can be used to compute 'expected' populations of cities in an urban hierarchy for comparison with observed populations, but is perhaps more effective as a graphical device, enabling departures from rank-size regularity to be ***seen*** when the rank-size distribution for an urban system is plotted together with that predicted by the rank-size rule. If wishing to compare empirical rank-size distributions, whether cross-nationally or over time, a difficulty arises with variations in the size of the largest city, but this is easily overcome by indexing the sizes of urban places in each hierarchy to a value of, say, 1,000 for the largest city.

Counterurbanization

It is appropriate to end this discussion of urbanization by noting that the process can occur in reverse. The phenomenon that has come to be known as ***counterurbanization*** was first noted in a study of early 1970s US data (Beale 1975), and has subsequently been observed in a range of more developed countries, sometimes to be supplanted again by conventional urbanization. As its name suggests it refers to a reversal of longstanding trends towards increased population concentration in urban areas (i.e., 'deconcentration') as the lifestyle and economic attractions of living in more rural surroundings have come to be appreciated and as technological advances have increased opportunities to work from home. It is not proposed to do more than mention the concept here, but it is a fascinating one that has rejuvenated some rural localities that had hitherto been stagnating because of long-term population drift to urban areas. Readers interested in pursuing it further might start by consulting Champion (1989).

Migration

In Chap. 1 the population balancing equation for an open population made explicit the fact that population change is the product of three processes: fertility (births), mortality (deaths) and migration (arrivals of in-migrants and departures of out-migrants). The first two of these processes were dealt with in Chaps. 6 and 4, respectively. We now turn to the third.

Mobility and Migration

Territorial, or spatial, mobility of population occurs at a variety of scales. At one extreme it is exceedingly temporary and involves movement over very short distances (e.g., a trip from one's home to the local shops and return); at the other it can entail movement halfway round the world, never to return (e.g., a permanent change of residence from England to Australia). ***Distance*** and ***permanence*** are two attributes that tend to make some territorial moves of greater interest than others. That is not to say that some moves that occur over relatively short distances and/or are distinctly temporary are not of interest. Local planning agencies, for example, often are intensely interested in ***journey to work*** patterns; the moves mostly occur over short distances and do not entail changes of place of residence, but their daily repetition has major implications for the provision of various services and facilities. Similarly a country, or a region, may have great interest in temporary population mobility that constitutes ***tourism***. Demographers, however, are most interested in territorial moves which are sufficiently sustained to be construed as ***changing a person's place of usual residence***, and it is moves of this type that qualify as ***migration***. This definition excludes certain categories of movement that sometimes are referred to as 'migration'; for example, nomadism, seasonal movement of persons who routinely live in two or more locations during the year, and movement of persons with no fixed place of residence. It is not, however, a clearcut definition in that there is no universally accepted way of determining when a change of place of usual residence has occurred. Sometimes an ***arbitrary minimum actual or intended duration of stay*** at a new locality is the criterion used, an approach which, when later behaviour fails to match stated intention when moving, generates a headache for analysts known as ***category jumping***. On other occasions respondents' ***perceptions*** of geographical moves as changing their own, or others', places of residence are accepted at face value. Where movement occurs across an international border the ***type of travel document*** a person carries may determine whether a change of place of residence is considered to have taken place (although people do not always behave in accordance with entitlements such documents may give them). Then again, some approaches to the study of migration focus not on individuals, but on ***net population aggregates***; if, after taking births and deaths into account, the population of an area has increased or decreased in size compared to some earlier point in time, a net rise or fall in the number of persons usually residing in that area due to migration is assumed to have occurred. It is argued as well that very short distance changes of place of residence (e.g., between apartments within the same building or houses in the same neighbourhood or town) should not be labelled 'migrations', and that migration is better defined as involving a change in ***locality*** of usual residence.

Data and Terminology

The term 'locality' arguably is less precise than 'place', but the intent in its use is clear, and in practice its meaning often is dictated by the units of area for which government statistical agencies provide data. In practice we rarely have data giving precise points of origin and destination for territorial moves, and therefore rarely have the capacity to classify moves with absolute precision by distance. Rather we have data indicating numbers of changes of residence ***between units of area***, and the larger those units are the greater the proportion of moves that are too short to cross a unit boundary and thus qualify as 'migrations' (in Australia, for example, there are fewer inter-State migratory moves than there are inter-SA1 migratory moves, because inter-SA1 moves ***within*** States are not inter-State moves). Idiosyncrasies also arise from the variable shapes of units of area, the closeness to unit boundaries of points of origin and destination, and the directions of moves from a point of origin. If area A is long and narrow, moves in the direction of the narrow dimension are more likely, and those in the direction of the long dimension are less likely, to qualify as migratory than those originating in area B, which is similar in area but roughly circular in shape. Second, the closer to an area boundary a point of origin is the shorter the move needed to cross that boundary and qualify as migratory. And finally, where a point of origin is not centrally located in an area, moves in certain directions qualify as migratory when moves of equal distance in other directions do not. These elements of imprecision and inequality usually have to be accepted as flaws in migration data about which nothing can be done.

Migration data normally are compiled with reference to a ***migration interval***. Migration intervals may be ***fixed*** (e.g., 1 year, 5 years, or an intercensal period) or ***indefinite*** (e.g., the lifetime of a population). The point is that to measure migration effectively requires a ***reference period*** within which movement has to have taken place. ***Lifetime migration*** is a concept sometimes studied when data relating to a fixed migration interval are not available. It uses data on places (or localities) of current residence and of birth, and defines a lifetime migrant as anyone for whom the two differ. It has very obvious flaws. Lifetime migration may be the net outcome of several separate migratory moves, and other things being equal, the older a person is the larger the number of individual moves that may be obscured. A person may even be a non-migrant in lifetime terms, yet actually have made several moves; it just happens that at the time data were gathered he/she had returned to live at his/her birthplace. Moreover, the timing of migration is unknown; even if the product of a single move, that move may have occurred anything from days before data were gathered to several decades previously.

Migrants are, of course, persons who engage in migration. They may do so more than once within a migration interval, so that if border crossings are being counted, the number of migrations typically exceeds the number of migrants, and by a greater amount the longer the migration interval. Other types of migration data, however, measure ***net*** movement of an individual over a migration interval (e.g., census data based on current locality of residence and locality of residence x years ago), in which case a count of migrants excludes those who moved away from, but later returned to, a locality of residence, but counts of migrants and of numbers of ***net*** migrations should be equal. ***Return migrants*** are persons who, having migrated, move back to the localities from which they migrated to live at a later date.

The terms ***origin*** and ***destination*** have already been used, and their meanings probably are clear. They are the locations (or localities, or areas) where migratory moves (or net moves) begin and end, respectively; the points/areas of ***departure*** and ***arrival***. A ***migration stream*** is the body of migrants who have common areas of origin and destination. Data for migratory moves among subareas of any defined larger area can be organized into a matrix of origin-destination pairs, each cell of which (except those on the leading diagonal, for which origin and destination are identical and therefore the frequency of migration is zero) corresponds to a migration stream. These streams form natural pairs in which origin for one is destination for the other and vice versa. The larger stream in any pair is often referred to as the ***stream***, and the smaller stream as the ***counterstream***, while the sum of stream plus counterstream is the ***gross migration interchange***, or ***gross migration*** (sometimes also referred to as the ***migration turnover***), for a pair of localities. The difference between flows in opposite directions between two localities, or between one locality and all, or a specified selection of, other localities, computed as arrivals minus departures, is ***net migration***. It may be positive, when arrivals exceed departures, or negative, when departures exceed arrivals, and in the context of any stream and counterstream is of identical magnitude, but opposite sign, at each of the two localities in question.

A fundamental distinction routinely made in the study of migration is between ***international migration*** and ***internal migration***. The former entails movement across national borders, and is distinctive because it normally is highly regulated by governments. This government involvement needs to be clearly understood in order to interpret migration flows. It also establishes ***administrative mechanisms*** (border control procedures and requirements for passports, visas, work permits, etc.) ***which can be important sources of data***, although these can be of varying degrees of completeness and reliability depending on the scope and incentive for avoiding the mechanisms and for behaving contrary to intentions expressed when interrogated by them. Other sources of data include aircraft and ship passenger manifests, census or survey data which establish residence in another country at birth or at a specified prior date, and occasionally population registers. Internal migration, in contrast, typically is a much freer category of population movement, with individuals and households responding to changing economic, social, environmental and life cycle circumstances as they see fit. A lack of regulation of movement generally means a lack of data comparable to border control data, and hence the need to rely on other data sources. Sometimes administrative systems which record changes of address for reasons other than keeping track of people's movements *per se* (e.g., in Australia, the medicare system) are potential sources of internal migration data. Whether they can be accessed for this purpose is, however, another matter, and ***population censuses are the major source of data on internal migration, followed by sample surveys and population registers***. The former two sources gather retrospective information on migration. Censuses do this either directly by asking questions about places of birth, residence at specific dates prior to the census, current residence, etc., or indirectly through techniques that employ survival ratios and data on age-sex composition and births (see Chap. 4). Sample surveys may also ask census-type questions, but in addition may gather individual migration histories, which date and give the origin and destination of each move. Population registers, on the other hand, record changes of place of residence at an individual level on an ongoing basis. Though of considerable potential, however, they are not widely available and their potential is largely unrealized.

There are exceptions to the generalization that internal migration is 'a much freer category of population movement', and they tend to be associated with totalitarian governments. In China, for example, the *hukou* system of household registration used to determine quite rigidly where one could live and work, and still has influence on, in particular, rural-urban migration. Under this system one is registered at one's *hukou* place and must apply to change that registration to obtain legal right of residency and, more importantly, access to state-subsidized welfare covering food, housing, education and permanent employment at a desired migration destination. The system is a *de facto* internal passport mechanism, and while approvals to change one's *hukou* place are easier to obtain in the post-economic reform era than they were in the pre-reform era, even today it can be difficult for rural peasants to gain approval to move to medium-sized and large cities.

The distinction between internal and international migration is matched by distinct terms for arrivals and departures. Internal arrivals are known as ***in-migrants*** and departures as ***out-migrants***, while the totalities of internal movements into and out of any given unit of area are known as ***in-migration*** and ***out-migration***, respectively. International arrivals and departures are referred to as ***immigrants*** and ***emigrants***, respectively, the totalities of movements into and out of a country being ***immigration*** and ***emigration***. It does happen that measures of migration for localities within a country combine (or are unable to separate) internal and international movement, or that a discussion of migration might be referring to ***either*** category of migration (e.g., the discussion of the I and O elements in the population balancing equation for an open population in Chap. 1). In such circumstances the terminology for internal migration prevails.

Demographic Perspectives on Migration

Donald Bogue (1993) has argued that demographers (and those with demographic interests in kindred disciplines – geography, economics and anthropology) study the territorial mobility of population from four perspectives. First, ***as a component of population change***, not only in the balancing equation sense of its being a factor in population growth or decline, but also in the sense of its altering population composition through selective movement by persons with particular characteristics. Second, ***as a mechanism for socioeconomic adaptation***; a means by which individuals, households and even entire communities adjust to changing social, economic, environmental and political realities which tend to 'push' them from one locality and 'pull' them towards another. Third, ***as a routine life-course event*** associated with such things as the attainment of adulthood (departure from the parental home to marry, attend institutions of higher learning, follow desired careers, or just establish independence), family formation (the desire to obtain housing perceived suited to the arrival, numbers and ages of children), employment transitions (job transfers, upward mobility in the labour market, and lapses into periods of unemployment), and retirement. Finally, demographers examine territorial mobility ***as rational entrepreneurship***. The focus here is on households or families as small businesses or enterprises seeking to maximise their wellbeing, an activity which may involve decisions to move or send individual members to other localities to improve the return on human and other capital.

Measuring International Migration

Administrative data on international migration are for many countries non-existent or of poor quality. There are also immense problems for international comparability of data in countries' different ways of defining and classifying migrants. Defining a migrant is less straightforward than defining a birth or a death, and even superficially similar approaches to the definition and classification exercise may mask subtle differences that result in quite varied estimates of the sizes of ostensibly the same migration flows. This was clearly demonstrated, for example, in an analysis of Australian and New Zealand data on migration between the two countries (Carmichael 1993), and these are countries with supposedly relatively high quality international migration data. One approach to filling gaps in administrative data on international migration, or to attempting to improve on data of suspect quality, is to seek out data gathered ***at the other end*** of migration streams involving the country of interest, although combining data from several sources obviously brings issues of comparability into sharp focus.

While in Chap. 1 we noted that ***absolute*** measures were of limited value in demography other than as input into ***relative*** measures, they perhaps have more of a place in analyses of international migration than in any other area of demographic analysis. There ***is*** interest in the sheer ***volumes*** of immigration, emigration and net migration flows, not only overall, but in particular streams and migration subcategories (family reunion, businesspersons, etc.).

Indirect Estimation of Net Migration

Where counts of immigrants and emigrants are not available, the population balancing equation in the form presented as Eq. 1.5 in Chap. 1 can be used to estimate net migration to a country between two censuses held at times 1 and 2. To recap:

$$M = (P_2 - P_1) - (B - D)$$

Where P_1 and P_2 = the total populations at times 1 and 2; B = births and D = deaths between times 1 and 2.

Estimates of intercensal net international migration ***by age and sex*** can be made using another methodology already discussed – forward survival. Known as the ***intercensal cohort-component method*** of estimating net migration, this was discussed for age groups at a second census which were already alive at an earlier first census in Chap. 4 (see Eqs. 4.75, 4.76, 4.77 and 4.78). The strategy is to forward survive age groups (or age ***cohorts*** – hence the technique's name) at a census held at time 1 to obtain their ***expected*** sizes assuming zero net migration at a second census held at time 2, n years later (when they form age groups n years older). Expected cohort sizes then are subtracted from those ***observed*** at the second census. Resulting differences are the cohort net population gains or losses due to migration over the intercensal period, and when the method is applied to national-level data these are net gains or losses due to ***international*** migration (applied at the subnational level net migration estimates produced are net gains or losses due to both international ***and*** internal migration). Recall that because survival chances between any two age groups of a given width differ for males and females, ***net migration estimates typically are computed separately for the two sexes***.

In outlining the intercensal cohort-component method in Chap. 4, separate equations were presented for (i) estimating expected cohort sizes at time 2 and (ii) subtracting these from observed cohort sizes to yield net migration. Separate pairs of equations also were presented to cover single-year and t-year age cohorts. This was done to break the procedure down and aid understanding of it, but it can be generalized in a single, much simpler looking equation as follows:

$$\mathrm{NM} = \mathrm{P}_{2,\mathrm{a}} - \mathrm{s}.\mathrm{P}_{1,\mathrm{a}-\mathrm{n}} \tag{7.17}$$

Where $P_{2,a}$ = the population in age group a, the age group defining the age cohort, at the second census (time 2); $P_{1,a-n}$ = the population in the corresponding n-year younger age group at the first census (time 1); n = the length of the intercensal period in years; s = the life table survival ratio from age group a-n to age group a. If age group a begins at exact age x and is of width t years, then this survival ratio will be given by $s = {}_tL_x / {}_tL_{x-n}$, these quantities being extracted from a relevant life table.

Equation 7.17 does not, however, deal with any younger age cohort at time 2 which was ***not alive*** at time 1, and thus is the product of ***births*** during the intercensal period. Net migration for any such cohort must be estimated by forward surviving the relevant number of ***births*** to obtain the expected size of the cohort at the second census date, then subtracting this expected size from the observed cohort size at the census. This procedure, too, was covered in Chap. 4 (Eqs. 4.79 and 4.80 in combination with Eqs. 4.76 and 4.78, respectively, for the single-year and t-year cohort situations), but again a single-equation summary is possible:

$$\mathrm{NM} = \mathrm{P}_{2,\mathrm{a}} - \mathrm{s}.\mathrm{B} \tag{7.18}$$

Where $P_{2,a}$ = the population in age group a, the age group defining the age cohort, at the second census (time 2); B = the number of live births during the period over which the cohort was born (which will be all or part of the intercensal period); s = the life table survival ratio from birth to age group a. If age group a begins at exact age x and is of width t years, then this survival ratio will be given by $s = {}_tL_x / (t \,.\, l_0)$, where ${}_tL_x$ and l_0 come from a relevant life table. The period over which the cohort was born will be the t-year period commenced the date of the second census in year y-x-t, where y is the year in which the second census was held.

The easiest way to carry out this second part of an intercensal cohort-component estimation of net migration (which again typically is conducted separately for males and females) is to define a single cohort covering all ages at the second census which are the product of intercensal births. That way B in Eq. 7.18 becomes all births (of a given sex) during the intercensal period. However, if the intercensal period is, say, of duration 10 years, and migration estimates based on Eq. 7.17 have been calculated for 5-year age cohorts, we may wish to treat those born during the intercensal period as two 5-year cohorts (aged 0-4 and 5–9 at the second census) instead of one 10-year cohort (aged 0-9).

The intercensal cohort-component method as outlined does tend to overstate or understate the implied number of deaths over an intercensal period, depending on whether a country, or a cohort, is an ‘emigration’ or an ‘immigration’ country or cohort. In the ‘emigration’ case the size of a cohort at the first census overstates, and in the ‘immigration’ case it understates, the average population at risk of dying during the intercensal period. Should this problem be judged to be severe one can obtain ***two*** estimates of net migration for each age cohort, one using standard forward survival from the first census (Eq. 7.17) or from intercensal births (Eq. 7.18), and the other using ***reverse*** survival from the second census, then ***average*** these estimates. The reverse survival companion to Eq. 7.17 is:

$$\mathrm{NM} = \mathrm{P}_{2,\mathrm{a}}/\mathrm{s} - \mathrm{P}_{1,\mathrm{a}-\mathrm{n}} \tag{7.19}$$

Where all elements have the same meaning as in Eq. 7.17.

The reverse survival companion to Eq. 7.18 is:

$$\mathrm{NM} = \mathrm{P}_{2,\mathrm{a}}/\mathrm{s} - \mathrm{B} \tag{7.20}$$

Where all elements have the same meaning as in Eq. 7.18.

Census-based studies of net international migration for countries which are net recipients of international migrants may also be carried out by focusing on ***the overseas-born***. Such studies do not, of course, reveal anything about migration of the native-born, but a modified version of the population balancing equation gives net intercensal migration of the overseas-born as:

$$\mathrm{M}_\mathrm{o} = \mathrm{P}_{\mathrm{o},2} - \mathrm{P}_{\mathrm{o},1} + \mathrm{D}_\mathrm{o} \tag{7.21}$$

Where $\mathrm{P}_{\mathrm{o},1}$ and $\mathrm{P}_{\mathrm{o},2}$ = the overseas-born populations at censuses held at times 1 and 2; D_o = deaths of overseas-born persons during the intercensal period.

There is no births element in this equation because, by definition, no births of overseas-born occur in a country. The equation is capable of being applied to overseas-born from particular source countries or regions as well as to total overseas-born, but does rely on having available data on deaths by birthplace. If not directly available, these can be estimated by applying total population age-sex-specific death rates at the midpoint of the intercensal period to means of the numbers of overseas-born in each sex-age group at the two censuses, summing to obtain an estimate of annual deaths, then multiplying by the length of the intercensal period in years. This procedure assumes similar mortality conditions among the total and overseas-born populations, and should not be used if the validity of such an assumption is in serious doubt.

Estimates of intercensal net migration of the overseas-born, or of subgroups thereof defined by birthplace, by age cohorts can also be made using the intercensal cohort-component procedures discussed above, assuming the availability of suitable life tables. 'P' quantities in the respective equations simply become numbers of overseas-born (or persons born in specified countries or regions) instead of numbers of total population. The interesting part of such a cohort-component exercise is that concerning migration in age cohorts born during the intercensal period. Superficially it might seem that no calculations are required; no persons in these age cohorts were native-born, and hence since all were born overseas during the intercensal period, all must have arrived as immigrants during that period. Net migration might seem to be simply the numbers of overseas-born enumerated at the second census, and certainly Eq. 7.18 would suggest this, since $B = 0$.

We have not, however, allowed that some immigrants during the intercensal period may have died before the second census. To take this possibility into account the reverse survival procedure using Eq. 7.20 can be used, with $B = 0$. This, however, effectively assumes that all mortality associated with immigrants born during the intercensal period occurred in the country of destination, which amounts to assuming that all migrated immediately they were born. This is not a plausible assumption; in reality migration will have occurred progressively between birth and the second census date, and in effect some who ***would*** have migrated will not have done so because they died in their countries of birth before being ***able*** to migrate.

One compromise would be to average the net migration estimate(s) yielded by Eq. 7.20 and the observed cohort size(s) at the second census, effectively assuming that half the mortality occurred after migration. But this is still likely to overestimate migration because childhood mortality is concentrated in the first year of life, and within that year very early in the first year of life. This makes it probable that well ***under*** half the mortality component added to net migration by reverse survival post-dated migration, with well ***over*** half of it ***preventing*** migration. In other words, the observed sizes at the second census of overseas-born cohorts born during the intercensal period might not be all that inaccurate as estimates of net migration after all! If precision is desirable, the distribution of mortality by age between birth and cohorts' ages at the second census should be taken into account in apportioning the mortality component added by reverse survival between that likely to have prevented, and that likely to have post-dated, migration (assuming migration to have been evenly distributed between birth and the second census date).

It is sometimes claimed that intercensal cohort-component procedures for age cohorts alive at both censuses can use ***census survival ratios*** rather than life table survival ratios. The former are calculated by dividing the number of persons in a sex-age (or age) group at a second census by the number of persons in the corresponding younger sex-age (age) group at an earlier census. For age cohorts born during the intercensal period, equivalent survival ratios divide numbers in an age cohort at the second census by the number of intercensal ***births*** over the period during which the age cohort was created. One advantage of this approach is claimed to be that intercensal differences in completeness of enumeration are incorporated in the survival ratios instead of contaminating net migration estimates. The method is of no use, however, where cohort-component procedures are applied to data for the total population, since census survival ratios assume a closed population and therefore yield zero net migration estimates. It may have some validity in application to data for ***the overseas-born***, in which case the census survival ratios would be based on data for the ***native***-born. The assumptions then are (i) that native-born mortality patterns apply to the overseas-born population (and to subgroups thereof defined by birthplace if they are the focus of analysis), and (ii) that the native-born population is closed, so that intercensal changes in the sizes of age cohorts are the product of mortality alone. Neither assumption may be especially accurate.

Measures Using Migration Flow Data

Few techniques have been specifically developed for analysing ***immigration and emigration flow data***. Several techniques are shared with studies of internal migration, and will be dealt with under that heading. Data on population flows across international borders should always be classified by type (temporariness or permanence) of movement, one of the functions of this classification being to (rather arbitrarily) distinguish immigrants and emigrants from other types of movers. Then, for the two migrant categories of movers information on countries of last and next 'permanent' residence, respectively, should be gathered to facilitate analysis of migration streams and counterstreams. Data on country of birth and citizenship may serve as substitutes for, or supplements to, such information. Border crossing databases can also usefully include information on reasons for moving (including data on visa categories for inward flows), and on various demographic, social and economic characteristics of movers in general and migrants in particular (since both international migration and non-migratory international movement tend to be selective by age, labour force characteristics, etc.).

Special difficulties attach to the construction of migration ***rates and ratios***, which often, though not always, make use of flow data. These difficulties emanate chiefly from the fact that each migration involves both a place of origin and a place of destination. While emigration (and out-migration) can be regarded as 'risks' for the populations of countries (or localities) of origin in the same sense as dying and giving birth are risks, immigration (and in-migration) are not risks for the populations of countries (or localities) of destination, since they involve people external to those countries/localities. Then again, besides being able to assume either positive or negative values, measures of net migration, as ***residuals*** after subtracting ***gross*** emigration from ***gross*** immigration, do not conform to the definition of the numerator of a demographic rate. Whether a person experiences the event of birth, death, marriage, emigration, immigration, etc. is clearcut (although in the case of immigration, as just noted, there is a problem defining an appropriate denominator for a demographic rate), but one cannot say that a person experiences the event of ***net*** emigration or ***net*** immigration.

Also an issue in the measurement of international migration flows is the fact that emigration and immigration are renewable events. Thus it is important to know whether numbers of ***migrations*** or of ***migrants*** are being counted. Typically, as noted earlier, the former are counted, and the obvious potential for double counting migrants in any migration interval poses further difficulty for notions of being able to calculate migration ***rates***.

Studies of international migration combine immigration (I), emigration (E), gross migration $(I + E)$ and net migration $(I - E)$ in a range of ratios that can reveal interesting features of and trends in migration flows. The ratio $\boldsymbol{E/I}$ is a measure of the extent to which immigration is counterbalanced by emigration; a value greater than 1.0 indicates net emigration, while the further below 1.0 the index gets the more dominant immigration is. Ratios of net immigration to total immigration (***(I − E)/I***; where $I > E$) and net emigration to total emigration (***(E − I)/E***; where $E > I$) give the proportions of total immigration and total emigration effectively added to and lost from a population in situations of net immigration and net emigration, respectively. Calculations of this type allow cognisance to be taken of the reality that similar ***net*** migration flows can be underpinned by very different ***total*** flows in the direction of net movement. Sometimes the bulk of a total flow is converted to net flow because there is little counterflow, whereas on other occasions a large counterflow means relatively little of a total flow becomes net flow. This type of measure can be very effectively applied with respect to individual migration streams as well as with respect to total flows into and out of a country.

The ratios ***I/(I + E)*** and ***E/(I + E)*** measure the relative contributions of immigration and emigration to gross migration. One of the more useful and widely used ratios, however, is the ***migration effectiveness ratio*** (MER), defined as the ratio of net to gross migration or:

$$\mathrm{MER} = [(\mathrm{I} - \mathrm{E}) / (\mathrm{I} + \mathrm{E})] \ . \ 100 \tag{7.22}$$

The MER measures the percentage of gross migration that is converted to population loss (a negative MER) or gain (a positive MER). Its value ranges from −100 where gross migration is entirely emigration to +100 where it is entirely immigration. Once again, it is not a measure applicable only to entire migration flows into and out of a country. It can be calculated for individual migration streams, and even for substreams defined by, for example, occupation.

While it was indicated above that several problems attend the construction of migration ***rates***, there are a number of measures in use in the study of international migration that are ***called*** 'rates', despite not satisfying the definition of such measures presented in Chap. 1. The ***crude rates of immigration, emigration, net migration and gross migration*** for a population are given, respectively, by:

$$\mathrm{CIR} = (\mathrm{I}/\mathrm{P}) \ . \ 1{,}000 \tag{7.23}$$

$$\mathrm{CER} = (\mathrm{E}/\mathrm{P}) \ . \ 1{,}000 \tag{7.24}$$

$$\mathrm{CNMR} = [(\mathrm{I} - \mathrm{E}) /\mathrm{P}] \ . \ 1{,}000 \tag{7.25}$$

$$\mathrm{CGMR} = [(\mathrm{I} + \mathrm{E}) /\mathrm{P}] \ . \ 1{,}000 \tag{7.26}$$

Where I = immigration during a calendar year; E = emigration during that year; P = the mean or mid-year total population in that year.

These are measures of the extent to which a country's population was augmented by immigration, depleted by emigration, and augmented or depleted by the net effect of both processes, and of the extent of international migration turnover in a population.

It is also possible to calculate a variety of ***specific*** 'rates' of international migration, the most common and useful probably being specific rates of ***net*** migration. For any population subgroup s (which might be an age group, a sex-age group, a birthplace group, etc.) the subgroup-specific rate of net migration is given by:

$$\mathrm{SSRNM} = [(\mathrm{I_s} - \mathrm{E_s})\,/\mathrm{P_s}]\ .\ 1{,}000 \qquad (7.27)$$

Where $\mathrm{I_s}$ = immigration into the subgroup during a calendar year; $\mathrm{E_s}$ = emigration from the subgroup during that year; $\mathrm{P_s}$ = the mean or mid-year population in the subgroup.

The 'rates' calculated using Eqs. 7.23, 7.24, 7.25, 7.26 and 7.27 have been presented as ***annual*** rates. If data on immigration and emigration are only available for some migration interval longer than a year, the same equations can be used, with the denominator P, or $\mathrm{P_s}$, becoming the relevant population at the ***mid-point*** of the migration interval. Thus we derive ***n-year*** rates, and if we wish to compare migration intervals all of which are of the same length no further modification is needed. If, however, migration intervals of varying length need to be compared, ***average annual*** rates can be computed by dividing n-year rates by n (although this procedure can be problematic where migration is cyclical and a shorter migration interval might include only an up-cycle or a down-cycle, while a longer interval might include both).

Equations 7.25 and 7.27 may also be used to calculate rates of net migration using ***census***-derived estimates of net migration. In this application the (I – E) numerators are obtained as single values using the population balancing equation or the intercensal cohort-component method. When Eq. 7.27 is applied to age cohorts using such data, it normally is applied only in respect of cohorts alive at both censuses defining the migration interval. Moreover, sometimes no attempt is made to obtain denominators which are mid-interval cohort sizes. Rather, the size of the cohort at the second census is used. Given that the 'rates' being calculated are not really rates at all, such latitude is not a major problem. The ***relative*** impacts of migration on different cohorts over a migration interval are adequately revealed.

International Migration and Population Growth

International migration is, of course, also of interest as ***a factor in national population growth***, as the population balancing equation made clear in Chap. 1. It was noted that total growth over any period ($\mathrm{P_2} - \mathrm{P_1}$) is the sum of net (international) migration (M) and natural increase (B – D). Various ratios of these quantities are commonly used in analyses of population growth to index the relative importance of migration and natural increase in growth. The percentages of total growth due to natural increase and to net migration are given by:

$$GNI = [(B - D) / (P_2 - P_1)] \ . \ 100 \tag{7.28}$$

And:

$$GNM = [M / (P_2 - P_1)] \ . \ 100 \tag{7.29}$$

And the ratio of net migration to natural increase, which we might call the ***components of growth ratio*** is given by:

$$CGR = [M / (B - D)] \ . \ 100 \tag{7.30}$$

Assuming positive natural increase, a positive value of the CGR indicates the percentage by which net migration augmented growth due to natural increase, while a negative value indicates the percentage by which it offset growth due to natural increase. If natural increase is negative (i.e., natural ***de***crease), as it has been in recent years in Japan and a number of European countries (e.g., in 2008–11, Belarus, Bosnia-Herzegovina, Bulgaria, Croatia, Estonia, Germany, Hungary, Italy, Latvia, Lithuania, Portugal, Moldova, Romania, the Russian Federation, Serbia and Ukraine), a positive CGR indicates the percentage by which net migration increased population loss due to natural decrease, and a negative one indicates the percentage by which it offset loss due to natural decrease.

Assessments of the contribution of migration to population growth often also seek to take account not only of migration *per se*, but of that element in natural increase that is attributable to migrants and perhaps also to their descendants. In the broadest view this type of assessment perhaps becomes rather ridiculous for a country like Australia, whose non-Indigenous population is ***entirely*** descended from migrants. At the other extreme the narrowest approach is to adjust net migration during a period for the natural increase for which migrants ***during that period*** are responsible. This can be done by forward surviving the population at a first census, plus births ***to that population*** during the intercensal period, to a second census date, then subtracting the resultant estimate of survivors from the second census population count. This becomes the migration contribution to intercensal growth, and the difference between it and total growth is the contribution of natural increase among the population present at the first census. The tricky part of this exercise is the estimation of intercensal births, which needs somehow to exclude births to migrant arrivals during the intercensal period. Otherwise the method becomes a straightforward series of intercensal cohort-component calculations.

In between these extremes is the type of exercise undertaken for Australia over many years by C.A. Price, who sought to monitor the contribution to population growth of ***post-Second World War migrants and their descendants***. Thus he took the 1947 Census population (itself with a significant immigrant, not to mention

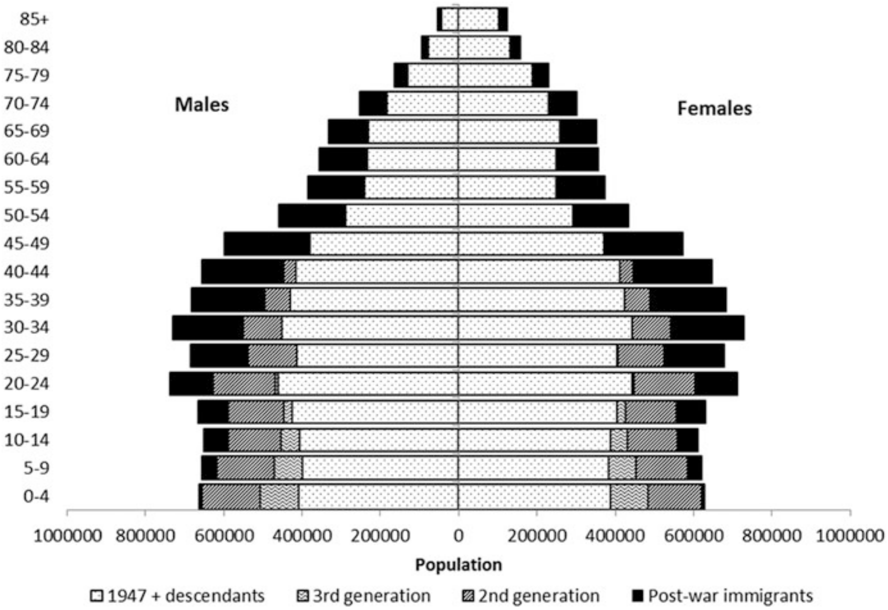


Fig. 7.2 Age-sex pyramid for Australia showing components of population added by post-Second World War immigration, 1993 (Source: Price 1996: 67–68)

immigrant-descended, element) as a base, thereafter progressively identifying, as they became features of the demographic landscape, (i) the post-war immigrant population, (ii) children of those immigrants (including half the children with one immigrant and one non-immigrant parent), and (iii) the third generation as components of the immigrant contribution to population growth. The methodology he used is complicated, summarized in Price (1996), and beyond being detailed here, but it produced the 1993 age pyramid shown in Fig. 7.2. You can probably imagine how, using data from which diagrams like Fig. 7.2 were constructed at regular intervals, population growth over those intervals could be apportioned by simple subtraction between that due to the 1947 Census population and its descendants, and that due to post-1947 immigrants and their descendants.

Measuring Internal Migration

Operationally, ***internal migration normally takes place when a person changes his/her place of residence from one statistical or civil division within a country to another***. This definition tends to presuppose censuses as the source of data on internal migration, and they are indeed the major source. They yield data on internal migration either through direct questions on place of residence, place of birth, duration of current residence, place of previous residence and place of residence at specified past dates, or indirectly through comparisons of population counts at successive censuses in conjunction with data on intercensal births and deaths, or through providing age distributions from which net migration estimates can be derived with the aid of survival ratios.

Census Migration Questions

It was noted earlier that ***place of birth*** data, in conjunction with ***place of current residence*** data, facilitate the study of ***lifetime migration***. The shortcomings of this concept also were discussed, but in the absence of better data, and ideally at a fairly broad level of geographic disaggregation, it provides some sort of indication of major population movements within and into (though not out of) a country over preceding decades. Internal lifetime migrants are those for whom place of birth is a locality within the country being studied which is different from the locality of current residence. International lifetime migrants are those whose birthplaces are external to the country being studied. With respect to internal movement, migration streams and counterstreams may be identified by focusing on place of birth/place of current residence pairs, always recognizing, of course, that at the individual level paths travelled between the two may have been indirect, and that migration patterns emerging may not be particularly representative of the ***recent*** past. One way of tackling the latter problem to a degree is to examine lifetime migration patterns for age cohorts. Very young cohorts could only have migrated relatively recently, and cohorts in the prime migration age groups (late teens to early thirties, say) are also worth focusing upon as likely to have migrated recently if they have migrated at all.

Place of birth data for a common set of spatial units from consecutive censuses can also be used to make indirect estimates of intercensal net migration for each unit. The relevant equation is:

$$\mathrm{NM} = (\mathrm{I}_2 - \mathrm{O}_2) - (\mathrm{s}_\mathrm{i}\mathrm{I}_1 - \mathrm{s}_\mathrm{o}\mathrm{O}_1) \qquad (7.31)$$

Where I_1 and I_2 = lifetime in-migrants to an area at the first and second censuses; O_1 and O_2 = lifetime out-migrants from that area at the two censuses; s_i and s_o = survival ratios applicable to the lifetime in-migrants and out-migrants, respectively, present at the first census over the intercensal period.

The main challenge in using Eq. 7.31 is to obtain satisfactory values for the two survival ratios. Data needed to calculate accurate values are rarely available, and so approximations of varying degrees of complexity are used. Perhaps the simplest approach is to set both survival ratios equal to the ratio total population aged n years and older at the second census divided by total population at the first census (or to the life table survival ratio $\mathrm{T}_\mathrm{n} / \mathrm{T}_0$ if an appropriate life table is available), where n is the length of the intercensal period. Other more elaborate procedures are outlined in United Nations (1970: 8–12).

Data on ***duration of current residence*** identify as migrants anyone for whom that duration is less than his/her current age. This approach has the capacity to identify as migrants persons who moved away from, then later returned to, their place (or locality) of birth, thereby providing a more accurate count of lifetime migrants. It also facilitates the identification of ***migration cohorts***, defined in terms of the timing of the most recent migratory move, and this allows more of a focus on the ***recent*** migration history of an area. These cohorts are, however, ***survivors*** of those who actually migrated during the periods defining them; survivors both in the sense of not having subsequently died and in the sense of not having subsequently migrated again. Other shortcomings are that, depending on the wording of the census question and the way respondents interpret it, moves which change a person's address without changing his/her ***locality*** of residence may be accorded the status of migrations; that there is no information on place, or locality of origin; and that therefore there is no capacity to focus on out-migration, net migration, or migration streams, or to distinguish internal from international migrants. Duration of residence data from successive censuses can, though, be used to examine ***re***migration, by surviving a duration of residence group from the first census to the second and comparing expected with observed survivors in the corresponding longer duration of residence group at the second census.

Many of the limitations for migration analysis of data on duration of current residence can be overcome if a question on ***most recent place of previous residence*** is also asked. This adds the missing information on localities of origin, and thus the ability to examine concepts listed in the previous paragraph. It also improves on place of birth data because it provides information on ***direct*** moves, rather than on moves which may be direct or very indirect. Analysed in isolation, however, most recent place of previous residence data share with place of birth data the problem that migrations may have taken place anything from within weeks of the census to many years previously.

In many respects the preferred census data for studying internal migration are data on ***place of residence at a specified past date***. Combined with data on current place of residence they provide a clear migration interval, often 1 year or 5 years, and where this coincides with the length of the intercensal period, successive censuses allow a continuous picture of internal migration to be built up. Census authorities also like the fact that one question classifies persons into migrants and non-migrants, provides a migration interval, and also yields the places of origin of migrants, in contrast to the two questions the duration of current residence/most recent place of previous residence approach requires. The method does revive problems associated with lifetime migration, in that if several moves take place within the migration interval only their net effect is measured, and if that net effect is a return to the place/locality of origin a person is deemed a non-migrant. But the strictly limited time frame within which these problems apply makes them far less serious. The capacity of respondents to recall accurately their places of residence at an arbitrary past date, as compared to their capacity to recall where they last moved from, has been questioned, and as with all census migration questions the tendency of persons to respond ***on behalf of*** others (e.g., a household head filling out census forms for all members of the household) can affect data quality through lack of knowledge of other household members' migration histories. But the place of residence at a specified past date approach does provide measures of in-, out- and net migration, for well-defined areas and migration streams, and over a well-defined migration interval.

Indirect Estimation of Net Migration

One method of indirectly estimating net migration already has been covered; that which uses Eq. 7.31 and is based on place of birth data for a common set of spatial units from consecutive censuses. Where reliable birth and death statistics are available for statistical or civil geographical divisions within a country on a place of residence basis, net migration for those divisions can also be estimated indirectly using the population balancing equation. Sometimes called the ***vital statistics method*** of estimating net migration, this approach can be applied either to the total population of a division, or to subpopulations defined by attributes which do not change (sex, birthplace, etc.) or which change predictably (age), provided births and deaths are available classified by those attributes. In the case of estimating net migration for an age cohort, the balancing equation $M = (P_2 - P_1) - (B - D)$ becomes, for an age cohort initially aged a, and an n-year intercensal period:

$$M_a = (P_{2,a+n} - P_{1,a}) + D_a \tag{7.32}$$

Where $P_{2,a+n}$ = the population in the age group n years older than age group a at time 2; $P_{1,a}$ = the population in age group a at time 1; D_a = deaths between times 1 and 2 in the cohort aged a at time 1. This number of deaths would need to be estimated from data on deaths by calendar year and single years of age using Lexis diagram principles.

Note that in Eq. 7.32 there is no reference to births (B), since intercensal births cannot possibly add to a cohort aged a (and therefore already born) at time 1.

It is important to appreciate that net migration estimates based on the population balancing equation are ***susceptible to error*** from two sources. First, if levels of underenumeration at the two censuses defining the migration interval are different, that difference is incorporated into the net migration estimate. So, for example, an improvement in census coverage (i.e., a reduction in underenumeration) is treated as a net migration gain. This can happen either overall, or in the context of individual age cohorts (e.g., as they age from their teens or early twenties, when their lifestyles might render them easily missed by census enumerators, to ages at which more settled lifestyles reduce the risk of this happening). The second source of error is the vital statistics used. In developing countries in particular, vital data often are of poor quality, and errors they contain are reflected in migration estimates (although to some extent undercounts of births and deaths may cancel out). Another point about net migration estimates calculated for subareas of a country using the vital statistics method was alluded to earlier; if a population is not closed, estimates reflect the combined effects of internal and international migration, and ***are not pure indicators of internal migration***. One partial way around this problem, assuming availability of the census and vital data by birthplace, is to apply the balancing equation to the native-born population only. This approach presupposes little international migration by the native-born, and if this is not the case the problem remains.

The intercensal cohort-component method of estimating net migration can also be applied in respect of statistical or civil divisions within a country to derive net migration estimates for those divisions. As when measuring international migration, two estimates can be made for any age cohort, one using forward survival and the other using reverse survival (Eqs. 7.17 and 7.19 for cohorts alive at the initial census, and Eqs. 7.18 and 7.20 for those born during the intercensal period). These approaches respectively assume that all migration occurred at the end of the migration interval, effectively failing to allow for deaths among migrants after arrival, and at the beginning of the migration interval, effectively overestimating such deaths, so that a common ploy is to ***average the results they yield***. This gives a result that assumes that migrations and deaths were evenly distributed over the migration interval, or alternatively that all migrations occurred in the middle of the interval. Thus, if NM_1 and NM_2 are the net migration estimates yielded for a cohort by forward survival (Eq. 7.17 or Eq. 7.18) and reverse survival (Eq. 7.19 or Eq. 7.20), respectively, we have:

$$\mathrm{NM} = (\mathrm{NM}_1 + \mathrm{NM}_2)/2 \tag{7.33}$$

We do not, however, need to evaluate NM_2 to obtain this average. It can be shown that:

$$\mathrm{NM} = [(1 + \mathrm{s})/2\mathrm{s}]\ .\ \mathrm{NM}_1 \tag{7.34}$$

Thus we can take a net migration estimate obtained by forward survival using Eq. 7.17 or Eq. 7.18 and adjust it using the survival ratio, s, from that equation to yield the average of forward and reverse survival estimates. Note that Eq. 7.34 works equally well whether the age cohort in question was alive at the first census (when Eq. 7.17 is used) or born during the intercensal period (when Eq. 7.18 is used). The survival ratio, s, must, however, be the one from whichever of Eq. 7.17 or Eq. 7.18 was used to find NM_1.

When the intercensal cohort-component method is applied to the estimation of net migration for statistical or civil divisions ***within*** a country, the use of ***census survival ratios*** (CSRs) as an alternative to, or in preference to, ***life table survival ratios*** (LTSRs) in Eqs. 7.17, 7.19 and 7.34 becomes more of an issue. Recall that a CSR is ***the number of persons in a sex-age (age) group at a second census divided by the number in the corresponding younger sex-age (age) group at an earlier census***. LTSRs for use with the intercensal cohort-component method can be obtained from male and female life tables for the area in question around the midpoint of the intercensal period (averages of LTSRs from tables for the endpoints of the period can also be used), or from appropriate model life tables. ***Their use is questionable, however, where census age data are of dubious quality***, and age distributions consequently are irregular; the combination of an irregular age distribution with a smooth set of LTSRs tends to produce distorted net migration estimates which do not sum over all areas within a country to zero. The latter situation is likely to stem partly from the population not being closed, so that net migration estimates incorporate international as well as internal migration, and it would be possible to smooth the census age distributions prior to application of cohort-component procedures. But ***the use of CSRs has the advantage that irregularities in age data and the effects of international migration tend to be absorbed into the survival ratios themselves***, rather than accruing entirely to the net migration estimates. This can result in CSRs which fluctuate rather than following the smooth trend with age characteristic of LTSRs. It can even result in CSRs above unity, which clearly are not accurate measures of survivorship. But the purpose is to measure net migration, not survivorship, and where internal migration is concerned, CSRs serve that purpose quite well.

Normal practice is to compute ***national*** CSRs, then use these in applying Eqs. 7.17, 7.19 and 7.34 to population data for subnational areas. One intuitively attractive feature of this approach is that, provided calculations are made for areas which together cover the entire country, net migration estimates for all areas sum to zero. ***The use of CSRs is not, however, without problems.*** First, although it has been stated that CSRs absorb the effects of international migration rather than have them affect the migration estimates themselves, this is only partly true. The CSR method works best for a closed population. While for an open population the effects of net international migration ***are*** absorbed into the CSRs, the subsequent application of these CSRs ***uniformly*** across a country assumes an ***even spatial impact*** of international migration, which is unlikely to coincide with reality. In a net immigration country, for example, areas (like major cities) which receive disproportionately large shares of net immigration are allocated ***less*** than their shares by this process, effectively converting some international migration gain to internal migration gain; other areas are allocated ***more*** than their shares of international migration gain, effectively lowering internal gain below its true level. One approach to overcoming this problem is to attempt to adjust the second census population for intercensal international migration, thus creating ***approximate intercensal closure*** before calculating CSRs. This, however, requires quality administrative data on international migration, which often are not available.

The second problem with using CSRs is the assumption that mortality conditions are uniform across statistical or civil divisions within a country. Especially in countries with high mortality this is unlikely to be the case, and some correction of net migration estimates for regional differences in mortality may be called for. Lack of data on such differences is often a barrier to making these corrections, but if one has a basic idea of regional variations in expectations of life at birth it is possible to use ratios of survival ratios in model life tables applicable to those regions to corresponding survival ratios in a model table applicable to the country as a whole as a basis for adjusting CSRs before application to regional population data.

A third problem with using CSRs to estimate internal migration via the intercensal cohort-component method is that the procedure assumes that, for any age cohort in any spatial unit, degrees of completeness of enumeration at the first and second censuses, relative to those applicable at the national level, are the same. Thus if an age cohort in an area is more underenumerated than the equivalent national cohort by, say, 2 % at the first census, it is assumed to remain 2 % more underenumerated than the national cohort at the second census as well. Checking the validity of this assumption normally is nigh impossible, but the mechanism that allows CSRs to inherently correct for census enumeration error is effective only to the extent that it holds. And even if it does hold, if there is overall underenumeration or overenumeration in the national population the total amount of net migration (over all ages) will be underestimated or overestimated accordingly. However, a ***rate*** of migration based on the population at the second census will be free of enumeration error, since both numerator and denominator will feature the same degrees of such error.

The CSR approach to the intercensal cohort-component method technically cannot deal with ***cohorts born during the intercensal period***, because, by definition, no cohort members were present at the first census, and therefore no CSR can be calculated for the cohort. However, if reliable birth registration data are available for the intercensal period, ***survival ratios analagous to CSRs can be calculated by taking the ratio of the size of the cohort at the second census to the number of births*** (of a given sex if calculations are being made for males and females separately) ***during the period over which the cohort was created***. These survival ratios can then be used in place of LTSRs in Eqs. 7.18, 7.20 and 7.34.

If reliable data on births are not available, approximations of net migration in 5-year age cohorts born during a 10-year intercensal period can be obtained using ***geographic area-specific child-woman ratios*** from the second census. If we define two child-woman ratios CWR(1) = children aged 0–4 divided by women aged 15–44 and CWR(2) = children aged 5–9 divided by women aged 20–49, then:

$$\mathrm{NM}_{(0-4)} = {}^{1}\!/_{4}\mathrm{CWR}(1)\ .\ \mathrm{NM}_{(f,15-44)} \tag{7.35}$$

$$\mathrm{NM}_{(5-9)} = {}^{3}\!/_{4}\mathrm{CWR}(2)\ .\ \mathrm{NM}_{(f,20-49)} \tag{7.36}$$

Where $\mathrm{NM}_{(0-4)}$ and $\mathrm{NM}_{(5-9)}$ are net migration for the cohorts aged 0–4 and 5–9 at the second census; $\mathrm{NM}_{(f,15-44)}$ and $\mathrm{NM}_{(f,20-49)}$ are net migration for the female cohorts aged 15–44 and 20–49 at the second census.

Choice of a method for estimating internal migration often is dictated by the availability of data. Where options exist, however, direct measures, such as can be obtained from data on place of residence at a fixed past date, or on duration of residence by place of last residence, normally are preferable. Among the three indirect approaches discussed, that using place of birth data (Eq. 7.31), the vital statistics method and the intercensal cohort-component method, it is impossible to be categorical as to which is preferable. Generally the intercensal cohort-component method, ideally executed using CSRs rather than LTSRs, seems to be favoured. One of its attributes is its inherent provision of age-specific detail, which the vital statistics method may, but does not necessarily, yield. But there are circumstances, often linked to the comparative quality of the data it and other approaches require, when the CSR-based cohort-component method is not the best option. These can be followed up in United Nations (1970: 35–36).

Migration Rates and Ratios

It has already been noted that the construction of migration rates and ratios is complicated by the reality that any migration involves both a place of origin and a place of destination. This difficulty can be set aside if one's focus is the ***totality*** of internal migration within a country or region during a migration interval. One's interest is only in ***whether*** individuals moved (internally) during the migration interval, not ***where*** they moved from or to. All members of the population are at risk of having become internal migrants, and a rate of migration can be calculated as follows:

$$\mathrm{M} = (\mathrm{m}/\mathrm{p})\ .\ 1{,}000 \tag{7.37}$$

Where m = the number of internal migrations or the number of persons migrating internally during the migration interval; p = the population exposed to the risk of migrating during the migration interval – i.e., the sum of migrants and non-migrants.

Whether the numerator in this measure is a count of ***migrations*** (which may be multiple for an individual) or of ***migrants*** will depend on the nature of available data and will determine whether M measures the incidence of ***moves*** or of ***persons moving at least once***. A count of migrants is perhaps preferable and tends to be the type of count yielded by the usual census approaches to obtaining internal migration data, subject to the limitation that migrants who had returned to their places of birth (the place of birth, place of current residence approach to lifetime migration) or to their places of residence at a specified past date cannot be detected. It is also arguable that p should exclude immigrants (i.e., persons who joined the population under study by ***external*** migration during the migration interval) because they were not at risk of internal migration throughout the migration interval, and any internal movements will have been suppressed in favour of their external movements in data based on place of birth or of residence at a fixed past date. The numerator m, too, should exclude external migrants, or their migrations (internal as well as external), as the case may be.

In the relatively rare event of data tabulating migrants by number of migrations being available for a migration interval, the possibility exists of computing not only ***both*** varieties of migration rate detailed above, but also rates of migrating ***only*** once (as opposed to ***at least*** once), more than once, and specific numbers of times in excess of once. In each case Eq. 7.37 could be used, with m redefined to be the number of migrants experiencing the relevant number of migrations. The denominator p would remain the sum of migrants and non-migrants, but could also be conceived of as the sum of migrants experiencing the relevant number of migrations, migrants experiencing some other number of migrations, and non-migrants. A further possible measure with data tabulating migrants by number of migrations is a rate of ***re***migration among migrants, defined as:

$$R = (m_{2+}/m)\ .\ 1{,}000 \tag{7.38}$$

Where $m =$ the number of persons migrating internally during the migration interval; $m_{2+} =$ the number of persons migrating internally on two or more occasions during the migration interval.

When analysis becomes focused on individual migration streams, strictly speaking only rates of out-migration can be calculated, since only persons at places of origin are at risk of becoming members of a stream (persons at places of destination are at risk of becoming members of the ***counter***stream, for which a rate of out-migration could also be calculated). Such rates of out-migration, which can be construed as measures of the attraction the place of destination holds for the population at origin, are given by:

$$M_{ij} = \left[m_{ij}/\left(p_i - m_{*i} + m_{i*}\right)\right]\ .\ 1{,}000 \tag{7.39}$$

Where $i =$ locality of origin; $j =$ locality of destination; $m_{ij} =$ migrations from i to j during a migration interval; $p_i =$ population at locality i at the end of the migration interval; $m_{*i} =$ in-migrations to locality i from all other localities during the migration interval; $m_{i*} =$ out-migrations from locality i to all other localities during the migration interval.

The denominator on the righthand side of this equation is a measure of ***the population alive and living in the locality of origin at the beginning of the migration interval who survived and were still living in the country/region for which internal migration is being studied at the end of the interval***. As they should be, immigrants (arrivals through external migration) during the migration interval are excluded (as one component of the m_{*i} that is subtracted from p_i). However, emigrants, who should have been ***in***cluded (as a component of the m_{i*} that is added to p_i) because, despite emigrating, they were also at risk of migrating between localities i and j are ***also*** excluded (because the census marking the end of the migration interval does not cover them).

When Eq. 7.39 is applied in respect of migration data based on residence at a specified prior date, its denominator is the population of the locality of origin at that date who survived the migration interval, less those who emigrated during the interval. When applied in respect of lifetime migration data, m_{ij} is persons born in locality i and resident in locality j at the census, and the denominator is the population born in locality i who survived to the census date, less those who were lifetime emigrants. In both cases m_{ij} are ***net*** moves and ***not necessarily direct*** moves. When applied, finally, in respect of data on duration of present residence and place of last previous residence, m_{ij} is persons whose ***most recent*** move within a selected migration interval (e.g., 5 or 10 years) was from locality i to locality j, and is a count of ***direct*** moves. The denominator, however, does not relate to a particular date and comprises persons at the census who had lived in locality i throughout the migration interval (persons enumerated in locality i (p_i) less those who had migrated there during the migration interval (m_{*i})), plus those whose ***most recent*** move within the migration interval originated in locality i (although they need not have lived there at the ***beginning*** of the migration interval).

While measures that express a migration stream m_{ij} as a ratio of the population at destination p_j are not migration rates in the sense of having 'at risk' denominators, they ***are*** calculated, and can be thought of as indexing the ***impact*** of particular migration streams on destination populations. One ***could*** adjust the denominator as before by subtracting all migrations during the migration interval with locality j as destination (m_{*j}) and adding all those with it as origin (m_{j*}), although because doing so does not yield a purer denominator of the 'at risk' type there is less necessity to do this. The case for adjustment is stronger if the aim is to compare the impacts of migration from each locality of origin on a range of localities of destination than if it is to compare the impacts on each locality of destination of migration from a range of localities of origin.

'Rates' of net migration and gross migration (migration turnover) within migration streams typically have the average of the adjusted census populations of the two localities as their denominator. Thus:

$$M_{ij-ji} = \left\{ \left(m_{ij} - m_{ji}\right) / {}^{1}\!/_{2} \left[\left(p_i - m_{*i} + m_{i*}\right) + \left(p_j - m_{*j} + m_{j*}\right) \right] \right\} \;.\; 1{,}000 \tag{7.40}$$

And:

$$M_{ij+ji} = \left\{ \left(m_{ij} + m_{ji}\right) / {}^{1}\!/_{2} \left[\left(p_i - m_{*i} + m_{i*}\right) + \left(p_j - m_{*j} + m_{j*}\right) \right] \right\} \;.\; 1{,}000 \tag{7.41}$$

Where i, j represent the two localities connected by a migration stream and counterstream; all elements otherwise have meanings the same as, or patterned on, those of elements in Eq. 7.39.

Besides focusing on migration streams and counterstreams connecting ***pairs*** of localities, analyses of internal migration may also examine in-migration to, out-migration from, and net migration between individual localities and ***all other localities***. Considerable debate has occurred over the most appropriate forms for the denominators of 'rates' of in-migration, out-migration and net migration calculated on this basis (see United Nations 1970: 41–42), but one view defines these rates for a locality i, respectively, as:

$$M_{*i} = (m_{*i}/p_i)\ .\ 1{,}000 \tag{7.42}$$

$$M_{i*} = [m_{i*}/\ (p_i - m_{*i} + m_{i*})]\ .\ 1{,}000 \tag{7.43}$$

$$M_{*i-i*} = \left[(m_{*i} - m_{i*})\ /\ \left(p_i - {}^1\!/_2\,(m_{*i} - m_{i*})\right)\right]\ .\ 1{,}000 \tag{7.44}$$

Where elements on the righthand sides of all three equations have the same meanings as in Eq. 7.39.

Other approaches argue that the denominator in Eq. 7.44 should also be used in Eqs. 7.42 and 7.43 (whence $M_{*i-i*} = M_{*i} - M_{i*}$), or that all three rates should use the non-migrant population ($p_i - M_{*i}$) as denominator.

The foregoing discussion assumes the availability of data on gross migration flows between origin-destination pairs. However, available migration data often consist of net migration estimates derived by indirect methods. Such data obviously lend themselves only to the calculation of ‘rates’ of ***net*** migration, and the question that arises is again one of the most appropriate denominator. Where net migration estimates are obtained by application of the vital statistics method (i.e., the population balancing equation) the appropriate denominator is usually taken to be the average of the sizes of the population of the locality (area) for which a net migration rate is being calculated at the two censuses defining the migration interval. Thus, where net migration for locality i is obtained by the vital statistics method, the rate of net migration is given by:

$$M_{*i-i*} = nm_i/{}^1\!/_2\,(p_{i,t} + p_{i,t+n})\ .\ 1{,}000 \tag{7.45}$$

Where nm_i = the net migration estimate for locality i; $p_{i,t}$ = the population of locality i at a census held at time t which marks the start of the migration interval; n = the length of the intercensal migration interval.

When net migration estimates for localities (areas) are obtained by applying the intercensal cohort-component method using census survival ratios it transpires that the same denominator is appropriate for a net migration ‘rate’ whether net migration is estimated by forward survival, reverse survival, or the average of the two. Adjustment factors applicable in the reverse survival and average cases are applicable equally to the ***numerators*** of rates, and thus conveniently cancel out. The relevant denominator is the equivalent of that in Eq. 7.44, except that only the net difference between flows out of and into a locality, not the separate flows (m_{*i} and m_{i*}), is known, and the equation for the rate of net migration takes the form:

$$M_{*i-i*} = nm_i/\left(p_i - {}^1\!/_2 nm_i\right)\ .\ 1{,}000 \tag{7.46}$$

Where nm_i = the net migration estimate for locality i; p_i = the population of locality i at the census marking the end of the migration interval.

The calculation, finally, of ***specific*** 'rates' of migration can generally be undertaken using whichever of the foregoing equations is appropriate given the nature of the migration data being used, with the various elements in the equation confined to members of the population subgroup specific to which the calculation is being made. Most commonly, as always, rates specific for age cohorts are of interest, and care must be taken to ensure that all migration flows, net migration estimates and population counts, as appropriate, pertain to the relevant cohort.

It should be kept in mind that internal migration 'rates' of the types just discussed pertain to particular migration intervals. There is obviously no problem with comparing rates calculated for different localities (defined at a given geographic scale) with respect to the same migration interval, or with comparing rates for a particular locality over different migration intervals ***provided these are of the same length***. But certain types of comparisons are not valid. You should not compare migration rates for 'localities' that are defined at different geographic scales. In the Australian context, for example, it makes no sense to compare a rate of internal migration calculated at the State/Territory level with one calculated at the SA1 level. Moves which are migrations at the latter level often will be too short to be migrations at the former level. Second, comparisons of migration rates based on migration intervals of different lengths are not valid. Some types of demographic comparisons where data are available for periods of differing length can be legitimized by calculating ***annual averages***; rates of population growth over intercensal periods of varying length are the obvious example. However, when dealing with migration data which are counts of migrants rather than of migrations, as the length of the migration interval increases so does the opportunity for individuals to have made multiple moves, and hence so does the degree to which a count of migrants underestimates the number of migrations. It follows that an annual average migration rate based on such data is biased downward to a greater extent the longer the migration interval, compromising the comparability desired. There is arguably less of a problem associated with annualizing rates of ***net*** migration, given that for any migration interval the net balances of migrants and of migrations are equal, but even then problems can arise, especially if age-specific rates are being computed. Assuming age cohorts are defined as at the end of migration intervals, where these intervals are of differing lengths, so are the ranges of ages over which cohort migration experience is being annually averaged. Bias is introduced, and annual averages are rendered non-comparable, because migration levels are highly variable by age. For example, the migration that is being averaged for a cohort aged 20–24 over a 5-year migration interval is migration between ages 15–19 and 20–24, a highly mobile phase of the life cycle. But that being averaged for a cohort aged 20–24 over a ***ten***-year migration interval is migration between ages 10–14 and 20–24, and the life cycle phase extending from ages 10–14 to ages 15–19 (when children are mostly still resident with their parents) tends to be a good deal less mobile than that extending from ages 15–19 to 20–24. In summary, the calculation of annual averages in an attempt to render comparable measures for intervals of differing lengths must be approached with caution, irrespective of the nature of the migration data.

Internal Migration Indices

Besides the various so-called ‘rates’ of internal migration (many of which are ratios, but not strictly rates – hence the title of the previous subsection) there are a number of other indices that can be calculated. A measure of the ***amount of population redistribution*** due to internal migration is the sum over the spatial units (localities) making up a country or region for which internal migration is being studied of all ***positive*** net migration estimates, or half the sum of the absolute values of ***all*** net migration estimates (since the sums of all positive and all negative estimates will be identical numerically). This can then be used to calculate a rate of population redistribution due to migration:

$$R_m = {}^1\!/_2 \Sigma_{i=1,k} |m_{*i} - m_{i*}| \,/\, {}^1\!/_2 (p_t + p_{t+n}) = \Sigma_{i=1,k} |m_{*i} - m_{i*}| \,/\, (p_t + p_{t+n}) \tag{7.47}$$

Where $k =$ the number of spatial units (localities) in the country or region; $p_t =$ its population at a census held at time t marking the start of the migration interval; $n =$ the length of the intercensal migration interval; m_{*i} and m_{i*} mean the same as in Eq. 7.39.

Note that a rate of population redistribution R_m is specific to the particular level of geographic disaggregation of the country or region on which net migration estimates used in its calculation are based, and its value will vary from level to level for a given country or region. Thus in the Australian context, for example, R_m values based on inter-State/Territory net migration estimates and on inter-SA1-level estimates would differ, the latter being higher because of the more intricate geographic disaggregation (and hence greater chance of a move crossing a boundary and becoming a migration) involved. This sensitivity of R_m to the level of geographic disaggregation makes its use for comparison inappropriate, except over time for a given country/region analysed at a given level of geographic disaggregation.

An advantage of migration data obtained from censuses is that because data on personal attributes gathered for the general population are available also for migrants, considerable scope exists for analysing ***migrant selectivity*** and ***migration differentials***. Studies of selectivity focus on whether members of certain subgroups of populations at localities of origin were more likely to become migrants than were members of others. Studies of migration differentials are concerned with whether in-migration rates differ among population subgroups at localities of destination.

Suppose that $m_1, m_2, \ldots, m_k$ represent the distribution of migrants at a locality of destination across the k categories of some characteristic, that $m = \sum m_i$, that $n_1, n_2, \ldots, n_k$ represent the distribution of non-migrants at the same locality across the same k categories, and that $n = \sum n_i$. Two indices of migration differentials (IMD) can be defined as:

$$\mathrm{IMD_i} = \{[(m_i/m) - (n_i/n)] / (n_i/n)\} \ . \ 100$$

And:

$$\mathrm{IMD_i} = \{[(m_i/n_i) - (m/n)] / (m/n)\} \ . \ 100$$

Both of which reduce to:

$$\mathrm{IMD_i} = [(m_i n - m n_i) / m n_i] \ . \ 100 \tag{7.48}$$

Positive values of $\mathrm{IMD_i}$ indicate an overrepresentation of migrants in category i; negative values indicate an underrepresentation. The numerical value of the index indicates ***the percentage by which*** migrants are overrepresented or underrepresented in category i compared to the representation of non-migrants in that category. Thus, for example, $\mathrm{IMD_i} = 29$ means that the concentration of migrants in category i is 29 % ***higher*** than the concentration of non-migrants in that category (concentration being measured by the percentage of total migrants or non-migrants who are in the category); and $\mathrm{IMD_i} = -55$ means that the concentration of migrants in category i is 55 % ***lower*** than the concentration of non-migrants in that category.

Indices of migrant selectivity can be constructed in a similar manner. However, whereas in measuring migration ***differentials*** the comparison that is made is between migrants and ***non-migrants*** at a place of ***destination*** (the latter representing the population being ***joined*** by the former), in measuring migration ***selectivity*** the comparison made is between migrants and the ***total population*** at a place of ***origin*** (since the latter represents the population being ***left*** by the former). Thus, suppose that $m_1, m_2, \ldots, m_k$ represent the distribution of migrants at a locality of origin across the k categories of some characteristic, that $m = \sum m_i$, that $p_1, p_2, \ldots, p_k$ represent the distribution of total population at the same locality across the same k categories, and that $p = \sum p_i$. Then the index of migrant selectivity (IMS) for any category i, following the pattern of Eq. 7.48, is given by:

$$\mathrm{IMS_i} = [(m_i p - m p_i) / m p_i] \ . \ 100 \tag{7.49}$$

Interpretation of values of $\mathrm{IMS_i}$ follows the same principles as interpretation of values of $\mathrm{IMD_i}$, except that the comparator is the concentration of total population (rather than non-migrants) in category i. Positive values indicate the percentage by which the concentration of migrants in category i ***exceeds*** that of the total population (positive selectivity) and negative values indicate the percentage by which the concentration of migrants in category i ***is below*** that of the total population (negative selectivity).

As well as being able to be calculated for individual localities (areas) of origin of internal migrants, indices of migrant selectivity can also be calculated for ***all*** internal migrants in a country or region (whence m_1, m_2, etc. are the sums of all internal migrants in categories 1, 2, etc. for the country or region and p_1, p_2, etc. are the total national or regional populations in those categories), for ***international migrants*** (whence m_1, m_2, etc. are the numbers of emigrants from a country in categories 1, 2, etc. and p_1, p_2, etc. are the total national populations in those categories), and even for ***international migrants to particular destinations*** (whence m_1, m_2, etc. are the numbers of emigrants from a country to a particular destination in categories 1, 2, etc. and p_1, p_2, etc. are again the total national populations in those categories).

In a similar manner indices of migration differentials can be computed for ***international migration*** (whence m_1, m_2, etc. are numbers of immigrants to a country in categories 1, 2, etc. during a migration interval and n_1, n_2, etc. are the populations in those categories who did not immigrate during the migration interval) and for ***international migration from particular countries of origin*** (whence m_1, m_2, etc. are the numbers of immigrants to a country from a particular origin in categories 1, 2, etc. during a migration interval and n_1, n_2, etc. are again the populations in those categories who did not immigrate during the migration interval).

Two other indices already dealt with in discussing population distribution and international migration, respectively, also have application in studies of internal migration (and in the former case in studies of international migration as well). The ***index of dissimilarity***, ID, (see Eq. 7.4) can be used to compare the percentage distribution of migrants across categories of any variable with an equivalent distribution of non-migrants or total population as appropriate. The distributions are more similar the closer ID is to 0, and more dissimilar the higher its value gets (the theoretical maximum being 100, but being rarely closely approached). Second, the ***migration effectiveness ratio***, MER, (see Eq. 7.22) can be applied to net and gross internal migration figures for individual localities within a country or region, or for locality pairs defining migration streams and counterstreams. In these applications the quantities I and E, instead of standing for immigrants and emigrants, stand for in-migrants from, and out-migrants to, ***all other*** localities in the country/region (effectiveness for a particular locality) or ***a particular*** locality (effectiveness for a particular migration stream).

Finally, if m represents the total number of internal migrants or migrations in a country or region during a migration interval, p_k represents the population of locality k and P represents the total national or regional population, then the proportions of migrants/migrations, m, expected to originate and terminate in locality k assuming equal likelihoods of movement from and to every locality are both given by p_k/P. It follows that the expected ***number*** of migrants/migrations from a locality i to a second locality j under the assumption that all localities are equally likely to provide migrants to and receive them from all other localities is given by:

$$m \,.\, (p_i/P) \,.\, \left(p_j/P\right)$$

The ***actual*** number of migrants/migrations from i to j (m_{ij}) can be compared with this ***expected*** number to yield an ***index of migration preference*** for moves from locality i to locality j:

$$\mathrm{IMP_{ij}} = \left[\mathrm{m_{ij}} / \left(\mathrm{m} \ . \ (\mathrm{p_i}/\mathrm{P}) \ . \ (\mathrm{p_j}/\mathrm{P})\right)\right] \ . \ 100 \tag{7.50}$$

The higher this index rises above 100, the more popular the two localities are as an origin-destination pair; the lower it falls below 100 the less popular they are. A value of 100 indicates that the scale of migration between the two localities exactly matches that within the country or region as a whole.

References

Beale, C. L. (1975). *The revival of population growth in non-metropolitan America* (Economic research service publication No. 605). Washington, DC: US Department of Agriculture.

Bogue, D. J. (1993). Spatial mobility and migration research: Editor's introduction. In D. J. Bogue, E. E. Arriarga, & D. L. Anderton (Eds.), *Readings in population research methodology* (Vol. 3, pp. 14-1–14-2). Chicago: Social Development Center for UNFPA.

Carmichael, G. A. (1993). Beware the passenger card! Australian and New Zealand data on population movement between the two countries. *International Migration Review, 27*(4), 819–849.

Champion, A. G. (1989). *Counterurbanization: The changing pace and nature of population deconcentration*. London: Edward Arnold.

Jones, H. R. (1990). *Population geography*. London/Thousand Oaks/New Delhi: Sage.

Newbold, K. B. (2010). *Population geography: Tools and issues*. Lanham: Rowman & Littlefield.

Price, C. A. (1996). *Immigration and ethnicity*. Canberra: Commonwealth Department of Immigration and Multicultural Affairs.

Statistics Division of the United Nations. (2008). *Principles and recommendations for population and housing censuses: Revision 2*. New York: United Nations.

United Nations. (1970). *Manual VI. Methods of measuring internal migration*. ST/SOA/SER.A/47. New York: United Nations. Available online at http://www.un.org/esa/population/techcoop/IntMig/manual6/manual6.html. Accessed 20th Feb 2014.

Chapter 8
Stable Population Theory

Stationary and Stable Populations

The concept of a ***stationary population*** was introduced in Chap. 4. The life table population, it was noted, was a stationary population. It was closed to migration, experienced a constant annual number of births (given by l_0), a constant and equal annual number of deaths ($\sum d_x = l_0$), thus had a constant size (T_0) and a zero growth rate, and had a constant age structure (given by the L_x-column).

A stationary population is a special case of a ***stable population***. Development of the stable population concept is, as was noted in Chap. 6, generally attributed to Alfred Lotka, who developed it in a series of papers published between 1907 and 1925, although the basic idea can be found in the work of Leonard Euler well over 100 years earlier. The discovery these men made was that ***if a closed population is subjected to constant schedules of fertility and mortality for a long period of time, eventually a fixed age structure develops which is independent of the age structure at the time the constant fertility and mortality schedules were first established***. Thus, a stable population can be defined as ***the limit to which a closed population's age (and sex) structure tends when it is subjected to constant age-specific schedules of fertility and mortality***. The 'stability' to which the concept refers is stability in age-sex structure, by which is meant an unchanging ***shape*** to the age-sex pyramid. It is perfectly possible for the ***size*** of a stable population, and therefore the ***numbers*** of people in each age-sex group, to increase or decrease. But once stability is attained the ***proportions*** of total population in each age-sex group ***do not change***. Moreover, the rate of growth of a stable population, r, usually referred to as the ***intrinsic rate of population growth*** or the ***intrinsic rate of natural increase*** (since stable populations are closed and thus migration is not a factor in their growth), is also constant or unchanging.

The intrinsic rate of natural increase provides the link between stationary and stable populations. ***A stationary population is a stable population whose intrinsic***

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_8

rate of natural increase is zero. In other words, the intrinsic birth and death rates (the constant crude birth and death rates of the stable population) are equal (yielding equal annual numbers of births and deaths).

The independence of the stable age-sex structures of stable populations from those that exist when constant schedules of fertility and mortality commence has been illustrated by Pollard et al. (1990), who take the very different age-sex pyramids for Sri Lanka and Sweden in 1960 (the former broad-based in the manner typical of young, high fertility populations and the latter much older) and show how they evolve subsequently if ***both*** populations are assumed to experience Sweden's 1960 fertility and mortality regimes indefinitely. By 2060 the two age-sex pyramids are virtually indistinguishable. The United Nations (1968: 6–8) presents similar evidence for East German and Thai pyramids projected under identical constant fertility and mortality schedules from the mid-1950s. The determinants of the shape of a stable population's age-sex pyramid are thus its constant fertility and mortality schedules alone, and have nothing to do with pyramid shape when those schedules first become established.

Because of the nature of the stability of stable populations described above, the intrinsic rate of natural increase, r, specifies not only the annual rate of population growth, but the annual rates of growth in numbers in each age group, and in numbers of both births (total and at each maternal age) and deaths (total and at each age).

Generating Stable Population Measures

For any population it is possible to define a ***stable equivalent*** population (age-sex) structure; the population structure that would eventually emerge ***if*** current fertility and mortality schedules were to prevail indefinitely. Conventionally a one-sex model is developed first, for females, since details like age-specific birth rates typically are only available for women and their reproductive ages are more well-defined. Parameters of the male component of the stable population then are defined in terms of those for the female component. The procedure requires, first, obtaining the relevant stationary population structure (the L_x-column of a life table based on current female age-specific death rates – see Chap. 4), then adjusting it so that the rate of growth is the intrinsic rate of natural increase r, not the zero growth rate of a stationary population.

The Intrinsic Rate of Natural Increase

It is important at the outset to differentiate the ***intrinsic*** rate of natural increase from the ***crude*** rate of natural increase. The latter applies to the existing population for which a stable equivalent is sought; the former to the stable equivalent, a hypothetical population that would evolve well into the future ***if*** fertility and mortality remained constant at current levels. Both populations have identical age-specific fertility and mortality schedules, but unless the original population is already stable, different age structures and thus different rates of natural increase. Unless a population is stable at the outset, the stabilising process alters its age structure, often radically.

The concept of a stable population is intimately related to the notion of reproductivity discussed in Chap. 6, and a reasonable approximation of the intrinsic rate of natural increase r can be obtained in terms of two measures introduced there – the net reproduction rate (NRR) and the mean length of generation (MLG). The NRR, the extent to which a cohort of women replaces itself with daughters if experiencing current age-specific female fertility and mortality rates, is a measure whose value is determined by those rates alone, and is unaffected by age composition. ***It therefore has the same value for both an observed population and that population's stable equivalent***, since the two by definition share the same age-specific female fertility and mortality schedules, and their invariably different age structures are not a factor (unlike the situation with their crude birth rates, which ***are*** affected by their different age structures). The NRR (denoted as R_0) thus can be thought of as a measure of ***the extent to which population increases in a generation***, and the MLG (denoted as T) as ***a measure of the period over which that increase occurs***. Then, making use of the exponential formula (on which the instantaneous rates of urban and rural population growth were based in Chap. 7 – see discussion preceding Eq. (7.7)), we have:

$$R_0 = e^{rT} \tag{8.1}$$

Whence $\log_e R_0 = rT$, and so:

$$r = \log_e R_0 / T \tag{8.2}$$

Derivation of a more precise estimate of r entails greater mathematical complexity than some readers will be comfortable with (Impagliazzo 1989), but leads along the way to another approximate equation for r whose evaluation requires only a slight extension of the procedure for obtaining the NRR and MLG presented in Chap. 6 (see Table 6.1). That equation is:

$$r = \left[-\alpha + \surd\left(\alpha^2 + 2\beta \log_e R_0\right)\right] / \beta \tag{8.3}$$

Where R_0 = the net reproduction rate as before; $\alpha = R_1/R_0$; $\beta = \alpha^2 - R_2/R_0 = (R_1/R_0)^2 - R_2/R_0$; R_1 and R_2 are what are known as the first and second ***moments*** of the curve representing the age schedule of net reproductivity.

Using the notation used in calculating the NRR and MLG in Chap. 6, assuming data for five-year age groups:

$$R_0 = \text{NRR} = \Sigma_{x=15-19,45-49}\left(b_{(f,x)}/F_x\right)\ \left({}_5L_i/l_0\right) \tag{8.4}$$

$$R_1 = \Sigma_{x=15-19,45-49} \left(b_{(f,x)}/F_x\right) \ ({}_5L_i/l_0) \ (m_x) \tag{8.5}$$

$$R_2 = \Sigma_{x=15-19,45-49} \left(b_{(f,x)}/F_x\right) \ ({}_5L_i/l_0) \ \left({m_x}^2\right) \tag{8.6}$$

Where m_x is the midpoint of age group x; other elements mean the same as in Eq. 6.11 in Chap. 6.

Besides R_0 being simply the NRR, or the sum of column (7) in Table 6.1, Chap. 6, R_1 is the numerator of Eq. 6.13 in Chap. 6 (the equation for the MLG), and is simply the sum of column (8) in Table 6.1. Furthermore, to obtain R_2 in Table 6.1 we would merely have to add a column (9) in which we would multiply values in column (8) by values in column (2). R_2 would be the sum of this column (9). Thus, while the evaluation of Eq. (8.3) above may seem complicated it really involves little more data manipulation than was required earlier to obtain the NRR and MLG.

If we were to add a column (9) to Table 6.1 we would obtain as its sum $R_2 = 888.57045$. Thus, with values of $R_0 = 0.92218$ (column (7)) and $R_1 = 28.10196$ (column (8)), we would have $\alpha = R_1/R_0 = 30.47$ and $\beta = \alpha^2 - R_2/R_0 = -34.93$, whence $r = -0.00265$.

Should data for single years of age be available, the equations for R_0, R_1 and R_2 become:

$$R_0 = NRR = \Sigma_{x=y,\omega} \left(b_{(f,x)}/F_x\right) \ (L_x/l_0) \tag{8.7}$$

Where all elements mean the same as in Eq. 6.9, Chap. 6.

$$R_1 = \Sigma_{x=y,\omega} \left(b_{(f,x)}/F_x\right) \ (L_x/l_0) \ (x + 0.5) \tag{8.8}$$

$$R_2 = \Sigma_{x=y,\omega} \left(b_{(f,x)}/F_x\right) \ (L_x/l_0) \ (x + 0.5)^2 \tag{8.9}$$

The Intrinsic Birth and Death Rates

The constant crude birth rate which will emerge once a population becomes stable is given, in terms of the intrinsic rate of natural increase r, by the following equations for five-year and single-year age group data respectively:

$$b = 1/\, \Sigma_{x=0-4,\omega} e^{-rm(x)}. \ ({}_5L_i/l_0) \tag{8.10}$$

Where ω = the oldest five-year age group represented in the population; m(x) = the midpoint of age group x; i = the lower limit of age group x.

And:

$$b = 1/\Sigma_{x=0,\omega} e^{-r(x+0.5)} . (L_x/l_0) \tag{8.11}$$

Where ω = the oldest single-year age group represented in the population.

Moreover, once the intrinsic birth rate b has been calculated, the intrinsic death rate d is simply:

$$d = b - r \tag{8.12}$$

The assumption at this stage is that, like earlier equations associated with the calculation of r, Eqs. 8.10 and 8.11 are evaluated using data for females only, giving rates of female births among the female population. However, whereas the value of r is the same for both females and males (as it must be for the sex ratio of the stable population to remain constant), and therefore also for the total population, the same is not necessarily the case with b and d. Values for the female, male and total populations may vary, provided that in each case $b - d = r$ (which does not vary).

To calculate b based on male data the numerator in Eq. 8.10 or 8.11 becomes the sex ratio at birth (s_b) instead of 1 (i.e., male births divided by female births; a value likely to be in the vicinity of 1.05), and the denominators are (i) evaluated using values of ${}_5L_i$ or L_x from the relevant ***male*** life table and (ii) each multiplied by s_b. To calculate b based on the total population the numerator becomes $1 + s_b$ and the denominators become the sums of the denominators of the separate equations for females and males. In each case Eq. 8.12 can then be used to obtain a value for d.

The Stable Age Distribution

The age structure of a stable population is a function of its intrinsic growth rate r and its intrinsic birth rate b. The constant proportion of the population in an age group x is given, for five-year and single-year age groups, respectively, by:

$$c\,(x) = b\ .\ e^{-rm(x)}\ .\ ({}_5L_i/l_0) \tag{8.13}$$

Where m(x) = the midpoint of age group x; i = the lower limit of age group x.

And:

$$c\,(x) = b\ .\ e^{-r(x+0.5)}\ .\ (L_x/l_0) \tag{8.14}$$

The life table quantities ${}_5L_i\,/\,l_0$ and $L_x\,/\,l_0$ in these two equations are sometimes generalized as a quantity p(x), meaning the average number of years spent in age group x by members of the life table stationary population.

You will note that Eqs. 8.10 and 8.11 feature the sums over ***all*** five-year and single-year age groups of values of $e^{-rm(x)}\ .\ ({}_5L_i\,/\,l_0)$ and $e^{-r(x+0.5)}\ .\ (L_x\,/\,l_0)$, respectively, while Eqs. 8.13 and 8.14 feature these same values ***for the particular***

age group for which c(x) is being calculated. Thus the normal approach to calculating c(x) values is to use the same calculating table as is generated to find b, so that effectively Eqs. 8.13 and 8.14 are rewritten with b replaced by the righthand sides of Eqs. 8.10 and 8.11, respectively. That is:

$$\mathrm{c\,(x)} = \left[\mathrm{e}^{-\mathrm{rm(x)}}.\ ({}_5\mathrm{L_i}/\mathrm{l_0})\right] / \left[\Sigma_{\mathrm{x}=0-4,\omega}\mathrm{e}^{-\mathrm{rm(x)}}.\ ({}_5\mathrm{L_i}/\mathrm{l_0})\right] \tag{8.15}$$

And:

$$\mathrm{c\,(x)} = \left[\mathrm{e}^{-\mathrm{r(x+0.5)}}.\ (\mathrm{L_x}/\mathrm{l_0})\right] / \left[\Sigma_{\mathrm{x}=0,\omega}\mathrm{e}^{-\mathrm{r(x+0.5)}}.\ (\mathrm{L_x}/\mathrm{l_0})\right] \tag{8.16}$$

It is also common to inflate values of c(x) by a factor of 1,000 or 10,000, so that stable age structures are given to either of those bases rather than to a base of 1.

Once again, the equations above are presented in the context of generating a one-sex, female, stable population. That is, life table functions come from the relevant female life table, b is the rate of female births among females, and c(x) is the proportion of females in age group x. The adjustments needed to generate stable male and total population age structures follow a similar pattern to those described above for generating values of b for stable male and total populations. In Eqs. 8.13 and 8.14 the b-elements obviously are adjusted as already described. Then, to generate a stable male age structure the life table function ${}_5\mathrm{L_i}$ or $\mathrm{L_x}$ is taken from the relevant male life table, and a multiplier of the sex ratio at birth, $\mathrm{s_b}$, is applied. To generate a stable total age (and sex) structure the life table function ${}_5\mathrm{L_i}$ or $\mathrm{L_x}$ is taken from the relevant female life table when calculating c(x) values for the female component of the population and from the relevant male life table when calculating them for the male component, with the multiplier $\mathrm{s_b}$ again being applied for males. In the total population case the modifications made in calculating b result in c(x) values that sum to 1.0 over all age groups for both sexes rather than for one sex.

Table 8.1 carries forward the example dealing with Australia in 2011 which was originally presented as Table 6.1 in Chap. 6 to illustrate calculation of the gross and net reproduction rates and the mean length of generation. It has already been shown that a minor extension of Table 6.1 yields a value of $\mathrm{r} = -0.00265$, and this value is used in the further calculations presented in Table 8.1, along with the observed 2011 value of $\mathrm{s_b} = 1.05712$. Calculations are based on five-year age group data.

The stable age distribution for females is given by column (6), and that for males by column (9). In each case c(x) values have been inflated to a base of 10,000 rather than 1.0. Columns (10) and (11) then give the stable age-***sex*** distribution (the stable age distribution for the ***total*** population), again to a base of 10,000 (the numbers of males and females in the stable age-sex distribution add to 10,000). You will note that despite the sex ratio at birth favouring males, the stable population features a slight excess of females. Their superior longevity more than compensates for the excess of males at birth. Males outnumber females through to age group 60–64, but females are decisively more numerous from ages 70–74 onwards.

Table 8.1 Generation of intrinsic birth rates and age distributions of stable female, male and total populations for Australia, 2011

Age group	m(x)	$e^{-rm(x)}$	Female ${}_5L_i/l_0$	(3) × (4)	(5)/$\sum$(5) (×10,000)	Male ${}_5L_i/l_0$ (× $s_b = 1.05712$)	(3) × (7)	(8)/$\sum$(8) (×10,000)	(5)/($\sum$(5) + $\sum$(8)) (×10,000)	(8)/($\sum$(5) + $\sum$(8)) (×10,000)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
0–4	2.5	1.006647	4.98215	5.015266	529	5.26177	5.296745	560	265	280
5–9	7.5	1.020074	4.97907	5.079019	536	5.25784	5.363386	567	268	283
10–14	12.5	1.033680	4.97708	5.144707	543	5.25541	5.432412	574	272	287
15–19	17.5	1.047467	4.97313	5.209190	550	5.24933	5.498500	581	275	290
20–24	22.5	1.061438	4.96666	5.271804	556	5.23470	5.556309	587	278	293
25–29	27.5	1.075596	4.96011	5.335075	563	5.21729	5.611696	593	282	296
30–34	32.5	1.089943	4.95153	5.396883	569	5.19683	5.664248	598	285	299
35–39	37.5	1.104480	4.93908	5.455117	576	5.17129	5.711586	603	288	301
40–44	42.5	1.119212	4.92031	5.506871	581	5.13746	5.749907	607	291	304
45–49	47.5	1.134140	4.89160	5.547761	585	5.08938	5.772069	611	293	305
50–54	52.5	1.149268	4.84883	5.572604	587	5.01784	5.766843	609	294	304
55–59	57.5	1.164597	4.78532	5.572969	588	4.91076	5.719056	604	294	302
60–64	62.5	1.180130	4.69046	5.535355	584	4.75095	5.606739	592	292	296
65–69	67.5	1.195871	4.54887	5.439863	574	4.51127	5.394897	570	287	285
70–74	72.5	1.211822	4.32732	5.243942	553	4.15216	5.031679	532	277	266
75–79	77.5	1.227985	3.96635	4.870620	514	3.60645	4.428667	468	257	233
80–84	82.5	1.244365	3.37541	4.200241	443	2.80275	3.487644	368	222	184
85–89	87.5	1.260962	2.44110	3.078135	325	1.77486	2.238031	236	162	118
90–94	92.5	1.277781	1.29832	1.658969	175	0.78985	1.009255	107	88	53
95–99	97.5	1.294824	0.41894	0.542454	57	0.20766	0.268883	28	29	14
100–104	102.5	1.312095	0.08550	0.112184	12	0.03387	0.044441	5	6	2
				$\sum = 94.789029$	$\sum = 10{,}000$		$\sum = 94.652993$	$\sum = 10{,}000$	$\sum = 5{,}005$	$\sum = 4{,}995$

$b_f = 1/\sum(5) = 0.01055$; $d_f = b_f - r = .01320$; $b_m = s_b/\sum(8) = 0.01117$; $d_m = b_m - r = 0.01382$; $b_t = (1 + s_b)/(\sum(5) + \sum(8)) = 0.01086$; $d_t = b_t - r = 0.01351$

Intrinsic birth and death rates for the female, male and total populations are computed at the bottom of Table 8.1. As foreshadowed earlier, values for the three populations differ.

Applications of Stable Population Theory

The primary purposes of this discussion have been to introduce the concept of a stable population and outline how its basic parameters can be calculated. While it is not proposed to develop applications of the theory at this juncture, it is appropriate to conclude with a brief statement indicating what the major applications are.

Stable population models have in the past been widely used in mathematical demography, demographic estimation and population projection. They can be used to estimate the age structures that will result from ongoing stability of fertility and mortality schedules in any nominated configuration. This sort of exercise has become increasingly relevant in the context of populations which have passed through the demographic transition and have established new, post-transitional fertility and mortality equilibria. It can also serve to illustrate what might be anticipated in a population if stability of fertility and mortality conditions in conformity with specified schedules were to be achieved.

Another application has been to the estimation of vital rates when an observed population can be regarded as approximating a stable population but has no, or no reliable, vital registration data. If relatively good census data exist for the population, an assumption that the population is closed is reasonable, and an estimate of the population growth rate is available, this information can be compared with sets of ***model stable populations***. The objective is to find the best fit, or a series of reasonable fits, then from the stable population(s) identified to obtain estimates of vital rates for the observed population.

This type of exercise has of course been of greatest interest among those studying the populations of developing countries, which were more likely to have approximated stability in the past than they have been more recently. Prior to major fertility transitions setting in, many of these populations had relatively stable (high) fertility in combination with declining mortality. Studies of such populations led to the development of correction factors which could be applied to the birth and death rates of stable populations, and to the labelling of these populations as ***quasi-stable***. Use of these correction factors extended the application of the general approach to demographic estimation to situations in which it would otherwise have been inappropriate. Discussions of fertility and mortality estimation using model stable age distributions can be found in United Nations (1968: Chapter VII, 1983: Chapter VII).

Stable populations are also useful for understanding population dynamics and their link to population structure. They can be used, for example, to demonstrate that fertility is a much stronger determinant of the proportions of total population under age 15 and over age 65 than is mortality. They are an experimental tool, allowing demographers to change parameters as they wish and to assess the consequences of those changes.

References

Impagliazzo, J. (1989). Stable population theory and applications. In S. A. Levin, T. G. Hallam, & L. J. Gross (Eds.), *Applied mathematical ecology* (pp. 408–427). Berlin/Heidelberg: Springer.

Pollard, A. H., Yusuf, F., & Pollard, G. N. (1990). *Demographic techniques* (3rd ed.). Sydney/Oxford/New York: Pergamon Press.

United Nations. (1968). *The concept of a stable population: Application to the study of populations of countries with incomplete demographic statistics*. ST/SOA/SER.A/39. New York: United Nations. Available online at http://www.un.org/en/development/desa/population/publications/mannual/model/stable-population.shtml. Accessed 27 July 2013.

United Nations. (1983). *Manual X. Indirect techniques for demographic estimation*. ST/ESA/SER.A/81. New York: United Nations. Available online at http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html. Accessed 26 July 2013.

Chapter 9
Population Projections

The Bread and Butter of Demography

Population projection was mentioned briefly in Chap. 4 when in Fig. 4.8 a Lexis-type diagram was presented to illustrate how projecting a population n-years into the future entailed (in part) forward survival of both the population at the beginning of the projection period and the births that would occur during the projection period. Population projection is arguably the most marketable skill demographers have – their bread and butter, widely assumed by potential employers to be their core business. It is fundamental to what is often termed ***'applied demography'*** (Rowland 2003), defined by Siegel (2001: 2) as 'the sub-field of demography concerned with the application of the materials and methods of demography to the analysis and solution of the problems of business, private non-profit organizations, and governments, at the local, national, and international levels, with a primary orientation toward particular areas and the present and future.' National statistical agencies, businesses and planning agencies at various levels from national through regional to local government recruit demographers first and foremost with an expectation that they will be skilled in preparing and/or making intelligent use of population projections. The provision of all manner of services and facilities is dependent on quality estimates of future demographic trends at all levels, from national to local, to ensure to the maximum extent possible that they are provided on time, in sufficient quantity and where, geographically, they are needed. Businesses also have a major interest in population projections as they plan the marketing of their products and the locations of their activities. And government policy formulation in areas like immigration, housing, education and ageing is underpinned by population projections.

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3_9

Population Projections and Population Forecasts

The terms 'projection' and 'forecast' are sometimes used interchangeably, and therefore loosely, by demographers to describe their efforts to predict future demographic conditions. It is, however, more correct to distinguish between the two. Until the mid-1940s official statistical agencies generally put out 'forecasts', although some called their numbers future population 'estimates', but the word 'forecast' was considered to imply a modicum of accuracy that caused U.S. agencies to be 'deeply embarrassed' when the unexpected upturn of the birth rate following the Second World War revealed 'a conspicuous divergence' between their population forecasts and subsequent reality (Keyfitz 1987: 242). Agencies could not, however, cease publishing the results of these activities, because despite the gross errors demand for them was strong. It was in this context, according to Keyfitz, that (1987:242) 'someone came up with a distinction between projections and forecasts. The former consisted of ***a non-committal working out of a set of stated assumptions and did not pretend to be an account of the future***. That would protect the agency from blame for the inevitable errors.' Thus, in the words of a Keyfitz section header, ***projections are marked by 'professional caution'***; associated ***'uncertainty'*** is routinely acknowledged (O'Neill et al. 2001: 217).

The standard approach to addressing this uncertainty is to produce ***a range of projections***. Particularly when preparing projections at national level or above (i.e., for major international regions or the world as a whole), demographers are apt to present ***series of alternative assumptions*** concerning the possible future courses of the components of population growth – fertility, mortality and migration – then undertake the mathematics that show the future consequences of different combinations of these assumptions. Commonly there is a medium, a high and a low variant for each component, although there may be more or fewer and the number of variants per component may differ. There may, for example, be more alternative fertility than mortality assumptions made because future fertility is deemed more unpredictable than future mortality. But projections show the size and usually also the composition of the population ***if*** particular assumptions were to hold true. They are 'conditional statements about the future' and are ***'nonjudgemental'*** (Smith et al. 2001: 3). The user is left to choose which projection seems most realistic for his or her purpose, or to contemplate a range of projections and the possibility that planning may need to be sufficiently flexible to allow for different plausible alternative eventualities. Projections create scenarios, which demographers usually do not claim to be forecasts. Instead they are, to again quote Keyfitz (1987: 246), 'ways of focusing discussion and judgement.' A developing alternative approach to dealing with uncertainty is ***probabilistic projections***. These use expert opinion, statistical analysis (the fitting of auto-regressive integrated moving average, or ARIMA, models) and/or analysis of errors in past projections to associate probabilities with alternative future regimes of fertility, mortality and migration (O'Neill et al. 2001; Wilson and Rees 2005).

The overriding message here is that ***from a predictive perspective projection is an inexact science***, and users need to be cognisant of that. Another indicator of this reality is ***the regularity with which statistical agencies update and revise their population projections***. Usually every new census, at least, is a signal to do this for national-level projections, establishing anew a (hopefully) robust base from which to project, but given that most countries do not have censuses as frequently as Australia does (five-yearly), other signals recommending revision may be the intercensal development of unanticipated trends in particular components of population change that call into question assumptions built into previous projections. International projections such as are prepared by agencies like the United Nations, on the other hand, are subject to constant updating as new census results for particular countries become available every year and things such as major epidemics emerge or are brought under a measure of control in different countries and regions. The classic recent example of a major epidemic has, of course, been HIV/AIDS, which suddenly dramatically impacted life expectancies in a range of countries after the mid-1980s and then, more recently, has seen its mortality impact in some of these countries rapidly reduced through the making available of anti-retroviral drugs.

One of the more widely used texts on demographic methodology in recent years defines population projections as ‘in general purely formal calculations, developing the implications of … assumptions that are made [about the future course of fertility, mortality and migration]’ (Preston et al. 2001:117). By contrast a population forecast is ‘a projection in which the assumptions are considered to yield a realistic picture of the probable future development of a population’ (Preston et al. 2001: 117). Or to quote Hinde (1998: 198), ‘The term *forecast* is used to indicate the actual predictions about which demographers feel reasonably confident.’ By these definitions ***the quality of a projection rests on its INTERNAL validity***; the mathematical soundness of the relations among demographic variables that it models. In the absence of computational error projections are always ‘correct’. ***The quality of a forecast***, in contrast, ***is dictated by its EXTERNAL validity***; or how well its predictions correspond with subsequent reality. In Smith et al.’s (2001: 3) words, ‘forecasts are ***explicitly judgemental***’. They can be proven right or wrong, or since absolute accuracy is nigh unattainable (whence forecasts are almost inevitably ‘wrong’ to some degree), relatively good or relatively poor, by subsequent events. Mention was made above of the practice of producing a range of projection variants. When this happens the tendency tends to be for a central, or medium variant to be regarded as the ‘best’ assessment of what the future will be like, and the one most likely to be elevated to the status of a ‘forecast’. Forecasting is by definition future-focused. So, too, generally is projection, but projection ***backwards*** into the past is also a possibility, although some would call such a process one of ***estimation*** rather than projection (George et al. 2004).

Some Other Features of Population Projections

Assuming a future focus, ***the predictive reliability of population projections that are accorded the status of forecasts clearly diminishes as the projection period lengthens***. Common sense says that regardless of the sophistication of one’s projection methodology, one has a better chance of being ‘ballpark accurate’ five years into the future than 20 or 50 years into the future. This is another reality demographers should constantly stress to users. The further into the future a projection extends, the less likely it is to merit being treated as producing a reliable forecast.

Projections produced by official statistical agencies have the advantage of being ***dispassionate***. That is, they are produced as ends in themselves by professionals skilled in such work, not with the aim of buttressing a particular business case or policy argument. Numbers produced will inevitably be inaccurate predictors of the future to a greater or lesser degree (the future is unknown and, in most respects, unknowable), but they are produced independently, not fabricated to support an argument. It cannot be claimed that projections produced by statistical agencies are beyond being contested, especially where forecast status is claimed for them. Demographers and statisticians may argue among themselves about the appropriate assumptions to make in carrying them out. International projections prepared by the International Institute for Applied Systems Research (IIASA), for example, are explicitly based on deliberations of expert groups convened to debate appropriate fertility, mortality and migration scenarios for major world regions (Lutz 1996). Planning agencies focused on particular geographic areas may, of course, be more intimately across demographic drivers in their areas of responsibility than their national statistical agencies, and may in consequence prefer to generate their own projections. However, one should be especially cautious when projections have been produced by interest groups or individuals pushing particular policy agendas. It is preferable that they emanate from an agency that is overtly indifferent to any policy agenda being spruiked, unless it can be convincingly argued that such independent projections are flawed and alternatives are warranted.

It is important to appreciate that population projection is usually not just about projecting population numbers. ***It is generally also about projecting population composition***. Most fundamentally it is about projecting ***age-sex composition***, but projections that explicitly focus on other compositional variables – e.g., ethnicity, household type and labour force status – are also undertaken and are receiving increasing attention in the literature (Wilson and Rees 2005).

Projections are carried out at ***a range of different geographic scales. International agencies*** such as the United Nations and World Bank prepare them for the world as a whole and for major, often continental or sub-continental, regions. For a comprehensive review of projections at this level and their history see O’Neill et al. (2001). ***National agencies*** naturally have a primary national focus, although they may also produce projections for major sub-national regions (e.g., Australia’s States and Territories) or populations (e.g., Indigenous Australians). Below national level ***planning agencies with responsibility for sub-national political units*** (e.g., states, provinces or cities), ***functional regions or local government areas*** also often produce projections of their own. These agencies are likely to be more knowledgeable about the forces for change operating within their jurisdictions, which at a local level might, for example, involve, in different areas, (i) conventional ageing of suburban populations, with housing initially occupied by young families, then children departing as they reach adolescence and adulthood followed by empty nesters ageing; (ii) constant turnover of similar types of people (e.g., in areas near universities with lots of student rental housing); or (iii) marked suburban compositional change as older residents relinquish housing stock and regeneration and perhaps also gentrification takes place. The latter occurs when those buying into a suburb are of higher socio-economic status than those who formerly lived there and upgrade housing stock. It is often sparked by urban expansion causing a premium to come to be placed on locational attributes of formerly lower class areas that were peripheral or otherwise less desirable when the city was smaller, but have acquired 'inner city' status as suburban sprawl has redefined the meaning of the term.

Projections at these different geographic levels raise different issues. Projection time horizons tend to shorten moving from international and national levels to local levels. Whereas at the former levels they typically extend several decades into the future and projections often only break population down by age and sex, small area projections usually have more immediate time horizons because of their short-term planning importance, a greater likelihood of taking other characteristics besides age and sex into consideration (e.g., education, labour force composition, rural-urban residence, household type), and the realization that forecast reliability diminishes rapidly as the projection horizon lengthens (O'Neill et al. 2001). At international level variability in the quality of data and missing data from certain countries can be major problems, requiring considerable adjustment and estimation. At a world scale, of course, in the absence of interplanetary travel, migration is not a potential source of population change and the focus is entirely on fertility and mortality. National-level projection options may benefit from greater data availability and reliability than exist for sub-national projections. Some data may only be available at national level; some may be available with greater frequency at that level; and if data are sample-based, sampling errors are likely to be smaller at national level than at sub-national levels. Moreover, ***migration tends to be a more major consideration in making sub-national, and especially local, population projections***. While there can be exceptions, international migration typically has a relatively modest, if at times temporally variable, impact on national population change, whereas ***internal migration is often the major source of sub-national population change***. It is apt to be ***more spatially variable and temporally volatile than either fertility or mortality***, making it difficult to project accurately and a major source of uncertainty in sub-national population projections. Especially at more local levels single occurrences, such as the closing of a major employer, a new residential development, or the launching of a major venture with a construction phase followed by an ongoing permanent boost to employment, can have significant migration-based effects on small area demographics, sometimes with limited warning. These sorts of issues make what are often termed ***small area population projections*** a specialized field of demographic endeavour, often inviting different choices concerning data, projection methodology and assumptions than might be made at other geographic levels (Wilson 2011, 2014). Choices may also be influenced by differing user needs. Commercial organizations, for example, may want a single 'most likely' short-term forecast refined by socioeconomic variables as well as age and sex to inform their marketing. Government planners may be interested in ageing and therefore in longer term projections that highlight the likely future health status and living arrangements of the elderly. Policy makers may prioritize alternative scenarios that attempt to model the demographic consequences of different policy options.

Projection of populations of subnational areas is not infrequently rendered more difficult than it would otherwise be by a ***lack of geographically consistent time series***. In other words ***subnational geographic boundaries*** other than basic ones that might divide a country into states, provinces etc. ***have a tendency to be redefined over time*** as, for example, cities expand into previously rural areas, perhaps capturing formerly isolated smaller settlements in the process. Boundary changes may require adjustments to historical data so as to approximate a current geographic configuration, and subnational projections may also need to anticipate future settlement expansion likely to generate further boundary revisions.

Population projections serve a range of different purposes. The most obvious one is the ***prediction of future population change***. Undertaken for this purpose it is imperative to play close attention to the plausibility of underlying assumptions. Without plausible assumptions, a projection will not yield a result meriting designation as a forecast. Second, projections can be used in a ‘What if?’ manner to study determinants of population change. In this application they have ***a simulation role*** that is ***illustrative*** of the effect of hypothetical changes, not predictive. Closely related to this second purpose is the third – ***presenting alternative scenarios*** as a way of trying to understand the range of possible future outcomes as components of demographic change are varied across plausible ranges. This type of exercise facilitates planning for worst-case outcomes. The fourth purpose is ***to support a particular political or economic agenda or to sound a warning*** about a perceived future threat, the aim in the latter instance being to stimulate preventive action. And the fifth purpose is ***to provide a rational basis for decision making.*** In this application a population projection with forecast status may be used as a base for projecting other phenomena – for example, the labour force (through application of appropriate labour force participation rates), housing demand (by incorporating assumed future household composition) or demand for educational services (by applying projected enrolment rates to relevant age groups).

George et al. (2004) observe that population projections take advantage of two strong features of demography – its ***accurate recording of demographic processes*** over lengthy periods and the ***momentum*** that links those processes for one time period with those for a later time period. Demographic futures, in the short to medium term at least, are usually intimately tied to demographic pasts, so projections based on past trends and relationships (p. 562) ‘often serve as forecasts of population change that are sufficiently accurate to support good decision making.’

Approaches to Population Projection

At one very basic level a distinction can be drawn between ***subjective*** and ***objective*** approaches to population projection. Subjective approaches lack clearly defined processes for analysing data; they rely on general impressions, intuition, analogy or even wild guesswork. The nature of the projection process is not clearly specified so as to be replicable by another analyst. Such approaches have their place when it comes to forecasting things like political change or technological change, but they are rarely if ever defensible ways of projecting demographic change.

Respectable approaches to population projection are objective. That is, the projection process is clearly specified in terms of its assumptions, data sources and the mathematical relationships employed. It is ***amenable to being replicated*** by someone else. Objective approaches do, though, have subjective elements – ***they involve judgement***, in respect of assumptions, data sources, key variables, appropriate time periods and functional forms. In addition it is important to appreciate that projection methodology is ***constantly evolving*** (Willekens 1990; Wilson and Rees 2005).

The selection of a projection methodology depends on ***the desired level of detail in the output*** (whether the aim is just to project total population or whether elements of population composition are also important) and ***the availability of requisite data***. Methodological sophistication is not necessarily advantageous if additional data it requires are of poor quality. George et al. (2004) list three broad categories of objective projection methods – ***trend extrapolation*** methods, ***the cohort-component method*** and ***structural models***.

Trend extrapolation methods are essentially what Hinde (1998) calls ***the mathematical method***. They fit mathematical functions to observed historical data and use these functions to extrapolate into the future. Such methods are typically used to project total populations, and while they may be separately applied independently to population subgroups, do not project population composition as such. They are mostly quick and simple to apply, and have minimal data requirements. They may thus be methods of choice when data series are incomplete, and/or time and/or budget are constrained. However, besides providing little or no information on the future demographic characteristics of a population these methods have a number of limitations. With the possible exception of logistic extrapolations, which imply a Malthusian population growth dynamic (S-shaped, with initial slow growth accelerating for a period then slowing again), they cannot be related to theories of population growth and so have limited utility for analysing the determinants of population growth. They make implicit assumptions about the continuity of population change according to the chosen mathematical model throughout the projection period, which can lead to unrealistic and even absurd results, even over relatively short time horizons. Their failure to explicitly factor in declining fertility rates has in recent times, for example, made them prone to overestimate future population. And when applied to population subgroups they are apt to ignore logical interdependencies among those subgroups, so that subgroup projections do not necessarily sum to an overall projection.

The cohort-component approach to projection is ***the most widely used approach for national and international projections***. Indeed Preston et al. (2001: 119) describe it as 'now nearly the only method used for [such] population projections, representing a rare consensus for the social sciences.' Broadly, it entails dividing a base population into subgroups assumed to be differentially exposed to the risks of fertility, mortality and migration, and separately estimating the changes over time for each subgroup. At a minimum these subgroups are defined by age and sex, but more complex divisions that also recognize variables such as race/ethnicity, rural-urban residence, country of birth (nationality), religion, educational attainment, etc. are also conceivable. Projection periods (the period between the date of the base population and the most distant date to which the projection is being carried out) are typically divided into intervals of length equal to the width of the age groups adopted, and projections are then carried out one interval at a time so that projected populations at dates ***within*** the projection period are also produced, not just an end-of-projection-period projected population. So, if single-year-of-age data are used, projected populations at single-year intervals are produced; if five-year age group data are used, projected populations at five-year intervals are generated (unless the methodology incorporates means of converting the five-year age group input data to single-year data, as, for example, the cohort-component projection procedure in the widely used United Nations MortPak suite of demographic computer programs does). The cohort-component method is traced back to Cannan (1895), was independently developed by Whelpton (1928, 1936), and was first used to project global population by Notestein (1945). It has become more detailed and sophisticated over time, not least as computers have eliminated the tedium formerly associated with its use, but its basic framework has changed little from that outlined by its pioneers.

Structural models typically come into play when planners and decision-makers encounter questions projection methods based entirely on demographic factors or the extrapolation of historical trends are ill-equipped to answer. These are questions concerning the demographic impact of major changes or developments in an area that will obviously have implications divergent from what would be expected in their absence – the opening of a large new industrial plant, or the building of a major new piece of transportation infrastructure, or the closing down of a major source of employment, for example. Structural models rely on relationships between demographic and non-demographic variables, basing projected population changes on projected changes in the non-demographic variables. These relationships are usually developed using regression-based techniques. Population projections carried out in this way allow for factors such as projected changes in the economy, land use, housing, the transport system and the environment. Two general categories have been recognized (Smith et al. 2001; George et al. 2004) – ***economic-demographic models*** and ***urban systems models***. The former are mostly used to project population and economic activities for nations and larger sub-national geographic areas; the latter are more a tool for small area projections at local geographic levels. Structural models may contain only a few equations and variables or may be extremely complex, with huge systems of simultaneous equations featuring numerous variables and parameters.

Trend Extrapolation Projections

Trend extrapolation projection methods fall into three subcategories – ***simple***, ***complex*** and ***ratio*** extrapolation methods.

Simple Extrapolation Methods

The simplest of the simple methods assumes ***linear change***; i.e., that annual ***absolute*** population change over a projection period will equal the mean annual change over the base period – the past period on which the projection is being based. The mean annual change over the base period is given by:

$$\Delta = (p_2 - p_1)/y \tag{9.1}$$

Where Δ is the mean annual population change over the base period; p_2 = the population at the end of the base period; p_1 = the population at the beginning of the base period; y = the length of the base period in years.

A projection assuming linear change beyond the base period is then generated using:

$$p_3 = p_2 + z\Delta \tag{9.2}$$

Where p_3 = the projected population at the end of a z-year projection period; p_2 is again the population at the end of the base period (i.e., at the beginning of the projection period); z = the length in years of the interval from the end of the base period to the date for which a projected population is required.

A second simple method of trend extrapolation assumes ***geometric change***; i.e., that the annual average ***percentage*** increase in population in future will be the same as the annual average change during the base period. The average geometric rate of population change during the base period is given by:

$$r = \left[(p_2/p_1)^{(1/y)}\right] - 1 \tag{9.3}$$

Where r is the average annual rate of population change over the base period; p_2 = the population at the end of the base period; p_1 = the population at the beginning of the base period; y = the length of the base period in years.

And a projection assuming geometric change beyond the base period is then generated by:

$$p_3 = p_2\,(1 + r)^z \tag{9.4}$$

Where aside from r, all elements have the same meaning as in Eq. 9.2.

The final simple method of trend extrapolation assumes ***exponential change***. This model is closely related to the geometric one, but views change as occurring continuously rather than over discrete intervals. The exponential rate of population change during the base period is given by:

$$r = [\ln(p_2/p_1)]/y \tag{9.5}$$

Where r is on this occasion the average annual exponential rate of population change over the base period; ln means 'natural logarithm of'; other elements have the same meaning as in Eq. 9.3.

So that a projection assuming exponential change in population is generated by:

$$p_3 = p_2 \, . \, e^{rz} \tag{9.6}$$

Where r is defined by Eq. 9.5 rather than Eq. 9.3; other elements have the same meanings as in Eq. 9.2.

Complex Extrapolation Methods

Complex trend extrapolation methods are distinguished from simple ones by the fact that they ***use base-period data for more than two dates***; that is, they don't just use data for the two endpoints of the base period. This makes them in theory better suited to dealing with non-linear population change, although there can be no guarantee that projections produced will be more accurate than would have been yielded by a simple extrapolation method. The first step in applying a complex extrapolation method is to assemble historical population data for different dates during the base period. These data must be based on consistently defined geographic boundaries for the area whose population is being projected. Next, parameters of the model selected to generate the projection must be estimated. And finally, the projection is generated using this model. It is important that consistent time units are used when (i) estimating complex extrapolation models and (ii) using them to project population values. Suppose the base period for a projection extended from 1991 until 2011. Time could be measured on a scale from 1991, 1992, …, 2011 or, equivalently, on a scale from 1, 2, …, 21. Either option could be selected, but once selected ***the same scale must be used when projecting into the future***. You cannot change scales beyond the date that separates the end of the base period from the projection period. If the base period is specified as 1991, 1992, …, 2011 the projection period must be specified as 2012, 2013, etc.; if it is specified as 1, 2, …, 21 the projection period must be specified as 22, 23, etc.

The simplest complex extrapolation method is again ***a linear model***, but one computed differently from the simple extrapolation method already outlined. It uses a linear regression equation, the parameters of which are first estimated using data for the base period then used to project the population linearly into the future. The equation takes the form:

$$p_i = a + bt_i \tag{9.7}$$

Where p_i is the population at time point t_i; a and b are the intercept and slope respectively of the linear regression line fitted to (p_i, t_i) pairs for values of t_i within the base period.

Having estimated a and b by ordinary least squares regression, Eq. 9.7 can then be used for projection purposes by substituting those values in conjunction with values of t_i that correspond to time points during the projection period.

A second complex extrapolation approach uses ***polynomial models***. These allow population change to be non-linear. The general form of a polynomial model relating population to time is given by:

$$p_i = a + b_1t_i + b_2{t_i}^2 + b_3{t_i}^3 + \cdots + b_n{t_i}^n \tag{9.8}$$

Where p_i is the population at time point t_i; a is the intercept term and $b_1, b_2, b_3, \ldots,$ b_n are the partial slope coefficients of an n-degree polynomial curve that best fits the (p_i, t_i) pairs for values of t_i within the base period.

The coefficients $a, b_1, b_2, b_3, \ldots, b_n$ can again be estimated using ordinary least squares regression techniques using data for the base period, then substituted in Eq. 9.8 along with values of t_i that correspond to time points during the projection period to generate the desired population projections. Equation 9.8 is a generalized equation that looks rather horrendous. In practice population projections normally make use only of ***second-degree*** polynomials, or ***quadratic functions***, whence:

$$p_i = a + b_1t_i + b_2{t_i}^2 \tag{9.9}$$

Quadratic functions can simulate a variety of population growth scenarios, including an increasing or a decreasing rate of annual growth, and population decline at an increasing or decreasing rate.

The third complex extrapolation approach models population change ***logistically***. This means a projection follows an S-shaped trajectory, with initial slow growth accelerating then tapering back again to approach zero as an upper limit on population size is approached. As previously noted, the logistic model is consistent with theories of constrained population growth, the best known of which is that of Malthus. Several equations are available to fit a logistic model to observed data for a base period, but the one presented here is a three-parameter model that can be fitted using NCSS (Number Cruncher Statistical System) software using the equation:

$$p_t = a/\left(1 + be^{-ct}\right) \tag{9.10}$$

Where p_t is the projected population at time t; e is the base of the natural logarithm; a, b and c are the three parameters estimated using the 'Three-parameter logistic' routine available under the 'Curve Fitting – General' suite of routines in NCSS, Version 9 (see http://www.ncss.com/).

The final complex extrapolation approach to population projection uses ***autoregressive integrated moving average (ARIMA)*** models, also referred to as ***Box-Jenkins*** models after Box and Jenkins (1970) who first systematically documented the approach. It is claimed to be appropriate for projection from base time series of medium to long length (i.e., at least 50 observations). Procedures used are complicated and no attempt will be made to outline them here. Mathematically adept readers are referred for more detail to Smith et al. (2001: 172–176) and George et al. (2004: 568–570), and further to Box and Jenkins (1970) (revised editions under the same title were published in 1976 and 1994, the latter with G.C. Reinsel as a third author) and McCleary and Hay (1980).

Ratio Extrapolation Methods

Ratio extrapolation methods are used where an area containing the population to be projected is ***a sub-area of a larger area for which projections already exist***. They are often used where geographic units exist at several levels such that those at each level aggregate to units at the next higher level, and ultimately all aggregate, with no omissions, to a single unit. This could be an entire country, a state or province, a city, etc.

The most straightforward ratio extrapolation method is the ***constant-share*** method. In this approach a smaller area's share of the larger area's population is held constant at a level observed during the base period – typically at the end of the base period where it transitions into the projection period. The relevant equation is:

$$p_{it} = \left(p_{ir}/p_{jr}\right) \cdot p_{jt} \tag{9.11}$$

Where i and j refer to the smaller sub-area for which a projection is required and the larger area within which it is located respectively; r and t refer to the reference year on which the projection is being based and the year for which the projection is required respectively; values of p are populations for the areas and time points that are defined by their subscripts.

Application of Eq. 9.11 requires data for only one historical date, so it is especially useful where changing geographical boundaries or poor records make constructing longer historical series difficult or impossible. The method's chief flaw is that it assumes all smaller areas grow at the same rate as the larger area within which they are located. This assumption will often not be plausible.

A second ratio extrapolation method is the ***shift-share*** method, which is designed to deal with changes in population shares. The literature offers several variants, the most straightforward of which extrapolates population shares linearly over time from a trend between end points of a base period. The relevant equation for this variant is:

$$p_{it} = p_{jt}\left[\left(p_{i2}/p_{j2}\right) + (z/y)\ \left(\left(p_{i2}/p_{j2}\right) - \left(p_{i1}/p_{j1}\right)\right)\right] \tag{9.12}$$

Where i and j refer to the smaller sub-area for which a projection is required and the larger area within which it is located respectively; subscripts 1 and 2 respectively denote the beginning and the end of a base period (the latter doubling as the beginning of the projection period); y and z denote the lengths of the base period and of the projection horizon at time t respectively; t refers to the year for which the projection is required; values of p are populations for the areas and time points that are defined by their subscripts.

The shift-share method should be used cautiously for longer projection horizons (e.g., 20 or 30 years). If, during the base period, sub-areas (i) grew very slowly or declined in population or (ii) grew very rapidly, these scenarios can respectively lead to substantial projected population losses or absurdly high projected population increases. These sorts of problems can be dealt with by building constraints into one's projections.

The final ratio extrapolation method is the ***share-of-growth*** method, which focuses on shares of population change rather than shares of population size. It assumes that a smaller area's share of population change in the larger area over the projection horizon will be the same as its share of change in the larger area during the base period. The relevant equation is:

$$p_{it} = p_{i2} + \left[\left(\left(p_{i2} - p_{i1}\right) / \left(p_{j2} - p_{j1}\right)\right)\ \left(p_{jt} - p_{j2}\right)\right] \tag{9.13}$$

Where i and j refer to the smaller sub-area for which a projection is required and the larger area within which it is located respectively; subscripts 1 and 2 respectively denote the beginning and the end of a base period; t refers to the year for which the projection is required; values of p are populations for the areas and time points defined by their subscripts.

This approach often yields more plausible projections than either of the other two ratio extrapolation methods, but does run into difficulty if a smaller area growth rate has the opposite sign from that of the larger area. There are ways of dealing with this, including setting an offending share to zero and not letting it change.

The Cohort-Component Method of Population Projection

If population projections are demographers' bread and butter the cohort-component approach to projection is their methodological bread and butter when carrying out projections. As already noted it is by some margin the most widely used approach to projection. Data availability permitting, it can be used at any geographic level, from the world and major regions thereof down through nations and then hierarchies of sub-national areas culminating in localized small areas, although small area projections needing to embrace local idiosyncrasies and data limitations frequently use other methods. The cohort-component approach derives much of its importance from the fact that projections of total population alone have limited utility. For broader planning purposes ***we are really interested in projecting not just total population, but population composition, or structure***. Education planners, for example, require projections of the numbers in age groups to which the services they oversee cater; planning for housing future population requires detailed projections by age, sex and marital status, among other information. The sorts of trend extrapolation methods detailed above can, as previously noted, be applied separately to structural subgroups of populations (defined by age and sex, and by other structural variables deemed relevant), but such an approach is cumbersome and has a serious logical flaw in that the separate projection of subgroups in isolation ignores interdependencies between them that cohort-component methodology inherently embraces. As a consequence subgroup projections tend not to sum to total population projections.

The name 'cohort-component' acknowledges that the method (i) segments a base population by at least age and sex, and projects age-sex groups (i.e., sex-specific ***cohorts***) individually, and (ii) projects for those cohorts the individual ***components*** of population change – the types of events through which population change occurs. When projecting a total population these components are ***fertility, mortality and net migration*** – people are added to the population by births, removed from it by deaths, and both added to it by in-migration and removed from it by out-migration, with the net impact of these two processes determining whether population increase or decrease from migration occurs. If the projection is focused on a subgroup of the population the number of components may be larger. For example, if the aim was to project the population of married females aged 30–34, births would not be a component of change (you can't be born married and aged 30–34). Deaths and in- and out-migration of married females aged 30–34 would be, but in addition we would need to consider (i) numbers of married women celebrating 30th or 35th birthdays, which would respectively see them join or exit the focal population, (ii) marriages of women aged 30–34, which would be a way of joining the focal population, and (iii) divorces of married women aged 30–34 and deaths of husbands of 30–34 year-old women, each of which would cause members of the focal population to cease being married and therefore to exit that population. Fortunately most population projections are projections of total population, so we will ignore this type of complication.

A cohort-component projection of population is what is known as a ***discrete-time*** model of population dynamics. This means that it only calculates population characteristics at certain moments of time separated by significant time intervals. Those intervals are ***as long as the age intervals employed to define cohorts are wide***. So if single-year age groups are employed the projection period is divided into one-year intervals and the projection is carried out one year at a time. If five-year age groups are employed the projection period is divided into five-year intervals and the projection is carried out five years at a time. These are not the only two options, but are by far the most common ones, with national projections for developed populations typically using single-year-of-age data to produce year by year projections into the future. Single-year age groups imply the availability of relevant ***complete*** life tables for males and females to use in one's projection. If only ***abridged*** life tables are available then five-year age groups and projection intervals will need to be used, unless a method for estimating single-year probabilities of dying from those for the wider age groups in the abridged life table is used (as, for example, in the MortPak projection routine marketed by the United Nations). Age group data, whether single-year or five-year age groups are employed, almost always terminate with an open age group – something like 70+, 80+, 85+ or 90+, depending on the level of survivorship to older ages in the population.

The Projection Process: A Simplified Model

The first step in a cohort-component projection is to ***select a base-year population structure***. This will often be a census-based structure, perhaps adjusted slightly to produce a mid-year structure if the census date deviated from mid-year. For simplicity we will assume a base population with an age-sex structure only (i.e., no additional compositional variables involved) and that the age groups are single-year groups. For now we will also assume a closed population – i.e., we will set aside the complication of migration and assume the only components of population change are mortality and fertility. This is the best way to grasp the logic of projection. Other complexities that make the projection process more realistic will be discussed subsequently. The projection is undertaken ***iteratively***; that is, sequentially over a series of separate short projection intervals (single-year intervals when using single-year age groups) until the date n-years into the future marking the desired outer limit or horizon of the projection (the end of the projection period) is reached. Along the way projections to intervening dates are also produced.

With the simplifying assumptions outlined, the second step involves ***projecting forward (surviving) the base population in each age-sex group to determine numbers surviving at the end of the first projection interval***, in this case one year beyond mid-year in the base year. This is a straightforward process using survival ratios derived from relevant male and female complete life tables (Chap. 4, Eq. 4.65). The equation, applied separately for males and females, is:

$$s_x = p_{x-1} \cdot (L_x/L_{x-1}) \tag{9.14}$$

Where s_x gives survivors aged x at the end of the projection interval; p_{x-1} is the base year population one year younger; L_x and L_{x-1} are taken from an appropriate complete life table for the sex (males or females) for which the calculation is being performed.

A modified equation is necessary to deal with survivors to the open age interval that single-year age distributions end with – age group x + (x and over, where x is usually 70, 80, 85 or 90). It is:

$$\mathrm{s_{x+}} = \mathrm{p_{(x-1)+}} \cdot \left(\mathrm{L_{x+}}/\mathrm{L_{(x-1)+}}\right) = \mathrm{p_{(x-1)+}} \cdot \left(\mathrm{T_x}/\mathrm{T_{(x-1)}}\right) \tag{9.15}$$

Where x is the exact age marking the lower bound of the open age interval; $\mathrm{L_{x+}}$ is the sum of $\mathrm{L_x}$ values from the relevant complete life table for all age groups making up the age interval x+ (which is the life table function $\mathrm{T_x}$).

Equations 9.14 and 9.15 produce sex-specific age distributions of survivors at ages 1 (last birthday) and older. ***Survivors aged 0 last birthday, however, have to be estimated separately***, because they were not alive at the date to which the base population pertains. They were born ***during*** the one-year projection interval separating the base-year population from the population being projected. Hence we need to estimate births during the projection interval, then survive those births to the end of the projection interval to obtain the missing survivors aged 0 last birthday at that time. Figure 9.1 provides a lexis diagram illustration of the projection of the population (of a given sex) aged 1 and older as survivors from the base-year population, but the population aged 0 last birthday as survivors from births (of that sex) during the projection interval (represented by the line AB).

The third step is therefore to ***estimate the number of births of each sex occurring during the projection interval***. We need births ***of each sex*** because we are projecting the male and female populations separately (since their mortality conditions, and also their migration patterns, which we are ignoring for the moment, generally differ), and so need to separately survive male and female births to the end of the projection interval to obtain male and female survivors aged 0 at that date. In theory, every birth is produced by two individuals, one of each sex, so that ideally births would be attributed to sexual unions, whose creation and dissolution would be treated explicitly in the projection framework. In practice, however, data to facilitate this are never available and the number of births is usually estimated by applying a relevant schedule of age-specific fertility rates to averages of the female populations in reproductive age groups at the beginning and end of the projection interval (as approximations of person-years lived by women at each reproductive age during the projection interval), then summing over all reproductive ages. Thus:

$$\mathrm{B_{t,t+1}} = \Sigma_{\mathrm{x}=15,49}\,\mathrm{F(x)} \;.\; {}^{1}\!/_{2}\,\left(\mathrm{p(f,x)_t} + \mathrm{p(f,x)_{t+1}}\right) \tag{9.16}$$

Where B is births; t and t + 1 are the dates defining the one-year projection interval; F(x) = the fertility rate for women aged x in the chosen fertility schedule (if rates are per 1,000, divide by 1,000 to get F(x)); p(f,x) = the female population aged x.

This equation can be rewritten as follows to express births as a function only of population at the beginning of the projection interval:

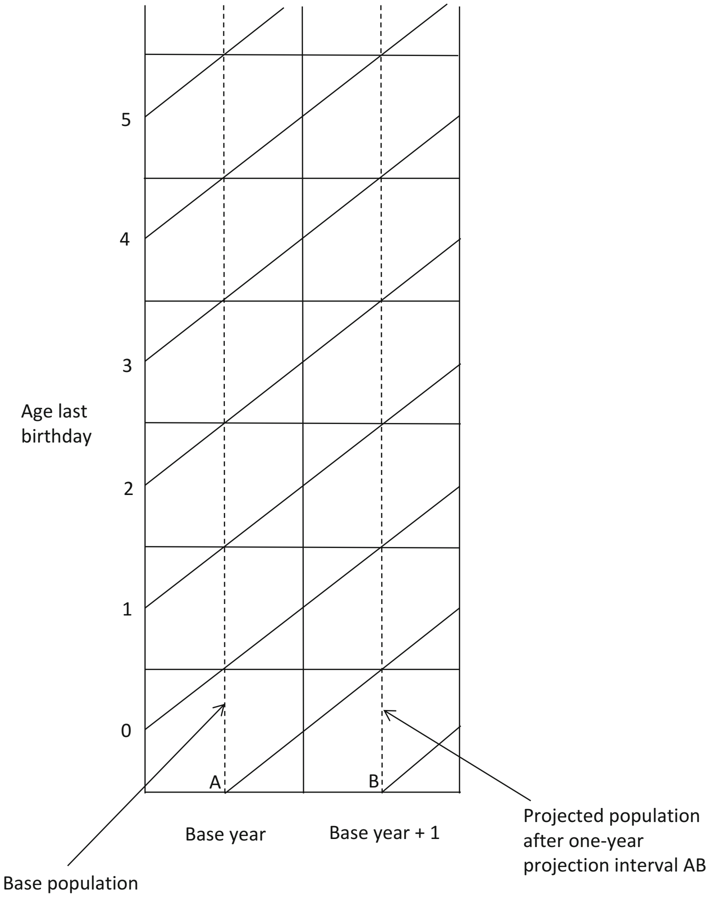


Fig. 9.1 Lexis diagram illustrating projection of a closed population over a one-year projection interval from a base year

$$B_{t,t+1} = \Sigma_{x=15,49}\, F\,(x)\ .\ {}^{1}\!/_{2}\ (p(f,x)_t + (p(f,x-1)_t\ .\ (L_x/L_{x-1}))) \qquad (9.17)$$

Where L_x derives from the complete life table summarizing the female mortality conditions assumed to prevail during the projection interval; other elements have the same meanings as in Eq. 9.16.

Total births during the projection interval $B_{t,t+1}$ are split into male and female births by applying an appropriate sex ratio at birth. This is likely to be around 105 males per 100 females (i.e., 105/100), but a more precise ratio based on recent actual experience of the population being projected is preferable if available. Thus we have:

$$B(f)_{t,t+1} = (1/\,(1 + SRB))\ .\ B_{t,t+1} \qquad (9.18)$$

And

$$B(m)_{t,t+1} = (SRB/\,(1 + SRB))\ .\ B_{t,t+1} \qquad (9.19)$$

Where $B(f)_{t,t+1}$ and $B(m)_{t,t+1}$ are female and male births, respectively, during the projection interval; SRB stands for 'sex ratio at birth'.

The fourth step in this simplified cohort-component projection model is to ***survive the births during the projection interval for each sex to the end of that projection interval***. This is accomplished by multiplying by L_0 / l_0 from relevant male and female life tables (Chap. 4, Eq. 4.67), and fills in the missing projected numbers of males and females surviving to age 0 at that date. Thus:

$$s_0 = B(s)_{t,t+1}\ .\ (L_0/l_0) \qquad (9.20)$$

Where the equation is applied separately for males and females; s_0 means survivors at age 0; $B(s)_{t,t+1}$ stands for births of sex s during the projection interval; life table functions come from the complete life table assumed to summarize mortality conditions for sex s.

Having projected the base-year population a year into the future, ***we can treat that projected population as a new base population*** and project ***it*** one year ahead in precisely the same way. Thus, in Fig. 9.1 the mid-year population in 'Base year + 1' is treated as the new base population from which to obtain a projected population for 'Base year + 2'. This process can be repeated for as many additional one-year projection intervals as we wish (or are game) to add to our overall projection, so that iteratively a five-year, 10-year or n-year projection is generated. In the n-year projection, the population aged n years and older will be survivors from the base-year population, while the younger population aged 0 to n−1 years will be survivors from birth cohorts born during the n-year projection period (Fig. 4.8 in Chap. 4).

The iterative process described does raise issues concerning the life tables and fertility schedules used. One option is to assume constant mortality and fertility throughout an n-year projection period, whence the life tables and fertility schedule don't change. However, if discernable trends in mortality and/or fertility are under way in years preceding one's base year, assumptions of constancy may be dubious. Other assumptions may be deemed more plausible, and mortality and/or fertility conditions may be allowed to change in clearly specified ways across the n-year projection period. Indeed, multiple alternative sets of assumptions may be adopted, with separate projections based on each set being prepared. This is what often happens when statistical agencies prepare national population projections that provide low, medium and high projection variants. What varies between them is the assumptions made about the future courses of fertility, mortality and also migration.

The general approach outlined above has assumed a base population structure with single-year age-sex groups. ***The same general approach is adopted when the base-year age-sex structure has five-year age groups***. The projection intervals are five years rather than one year long; the iterative process produces intermediate projections at five-year rather than single-year intervals; and the n-year projection period is such that n is a multiple of 5.

Figure 9.2, modelled on Fig. 9.1, provides a lexis diagram illustration of the projection of the population (of a given sex) aged 5 and older as survivors from a base-year population, but the population aged 0–4 as survivors from births (of that sex) during the five-year projection interval (again represented by the line AB). Having chosen a base-year population structure the second step involves ***projecting forward (surviving) the base population in each five-year age-sex group to determine numbers surviving at the end of the first projection interval***, this time five years beyond mid-year in the base year. This again is a straightforward process using survival ratios derived from relevant male and female abridged life tables (Chap. 4, Eq. 4.66). The equation, once again applied separately for males and females and using a standard life table type of notation, is:

$${}_5s_x = {}_5p_{x-5} \cdot ({}_5L_x / {}_5L_{x-5}) \tag{9.21}$$

Where ${}_5s_x$ gives survivors at the end of the projection interval in the five-year age group commencing at exact age x (x = 5, 10, 15, etc.); ${}_5p_{x-5}$ is the base-year population in the next younger five-year age group; ${}_5L_x$ and ${}_5L_{x-5}$ are taken from an appropriate abridged life table for the sex (males or females) for which the calculation is being performed.

As before, a modified equation is necessary to deal with survivors to the open age interval that five-year age distributions often end with – age group x+ (x and over). It is:

$$s_{x+} = p_{(x-5)+} \cdot \left(L_{x+} / L_{(x-5)+}\right) = p_{(x-5)+} \cdot \left(T_x / T_{(x-5)}\right) \tag{9.22}$$

Where x is the exact age marking the lower bound of the open age interval; L_{x+} is the sum of L_x values from the relevant abridged life table for all age groups making up the age interval x+ (which is the life table function T_x).

Equations 9.21 and 9.22 produce sex-specific age distributions of survivors at ages 5–9 (last birthday) and older. ***Survivors aged 0–4 last birthday, however, have to be estimated separately***, because once again they were not alive at the date to which the base population pertains. They were born ***during*** the five-year projection interval separating the base-year population from the population being projected. Hence we once more need to estimate births during the projection interval, then survive those births to the end of that interval to obtain the missing survivors aged 0–4 last birthday at that time. Figure 9.2 provides a lexis diagram illustration of the projection of the population (of a given sex) aged 5 and older as survivors from the base-

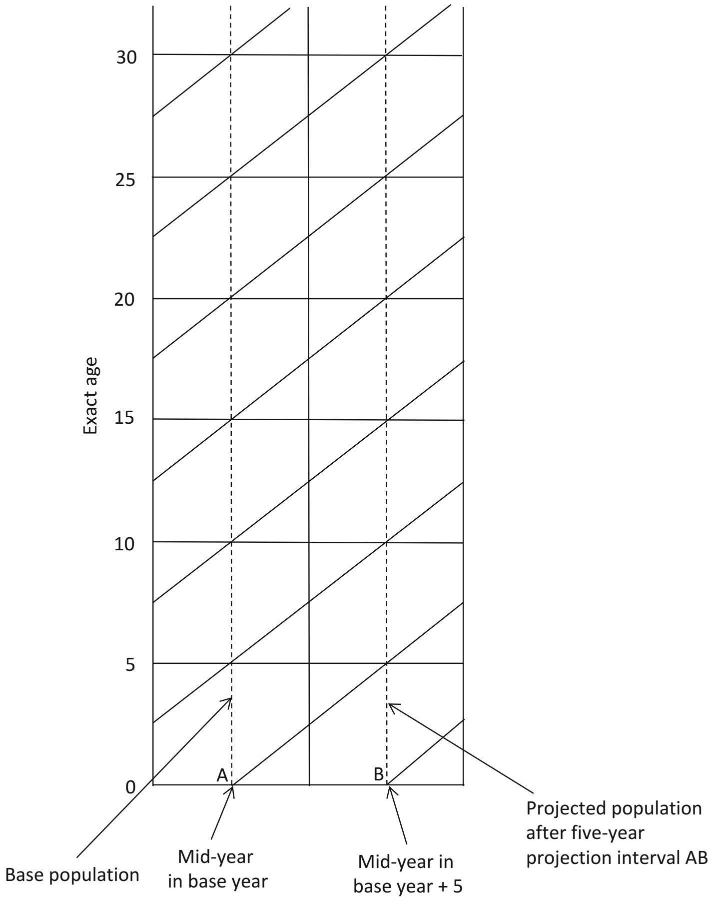


Fig. 9.2 Lexis diagram illustrating projection of a closed population over a five-year projection interval from a base year

year population, but the population aged 0–4 last birthday as survivors from births (of that sex) during the five-year projection interval (represented by the line AB).

The third step is therefore again to ***estimate the number of births of each sex occurring during the projection interval***. Once more we first estimate total births of ***both*** sexes by applying a relevant schedule of age-specific fertility rates to averages of the female populations in reproductive age groups at the beginning and end of the projection interval (as approximations, when multiplied by 5, of person-years lived by women at each reproductive age during the five-year projection interval), then summing over all reproductive ages. On this occasion we utilize data for five-year age groups rather than single-year age groups. Thus:

$$B_{t,t+5} = \Sigma_{x=15-19,45-49} F(x) \ . \ 5 \ . \ {}^{1}\!/_{2} \left(p(f,x)_t + p(f,x)_{t+5}\right) \qquad (9.23)$$

Where B is births; t and t + 5 are the dates marking the beginning and end of the five-year projection interval; x is a five-year reproductive age group; F(x) = the fertility rate for women aged x in the chosen fertility schedule (if rates are per 1,000, divide by 1,000 to get F(x)); p(f,x) = the female population aged x.

As before with Eq. 9.16, this equation can be rewritten to express births as a function only of population at the beginning of the projection interval:

$$B_{t,t+5} = \Sigma_{x=15-19,45-49}\, F(x)\ .\ 5\ .\ {}^{1}\!/_{2}\ (p(f,x)_t + (p(f,x-1)_t\ .\ ({}_5L_x/{}_5L_{x-5}))) \tag{9.24}$$

Where x−1 means the next younger five-year age group to age group x (so if x is 15−19, x−1 is 10−14); ${}_5L_x$ derives from the abridged life table summarizing the female mortality conditions assumed to prevail during the projection interval; other elements have the same meanings as in Eq. 9.23.

Total births during the projection interval $B_{t,t+5}$ are split into male and female births by applying an appropriate sex ratio at birth. A precise figure based on recent actual experience of the population being projected is preferable, but in its absence a ratio of 105 males per 100 females (105/100) is a reasonable approximation for most populations. Thus we have:

$$B(f)_{t,t+5} = (1/\,(1 + SRB))\ .\ B_{t,t+5} \tag{9.25}$$

And

$$B(m)_{t,t+5} = (SRB/\,(1 + SRB))\ .\ B_{t,t+5} \tag{9.26}$$

Where $B(f)_{t,t+5}$ and $B(m)_{t,t+5}$ are female and male births, respectively, during the projection interval; SRB stands for ‘sex ratio at birth’.

The fourth step in this simplified cohort-component projection model based on five-year age groups is to ***survive the births during the projection interval for each sex to the end of that projection interval***. This is accomplished by multiplying by ${}_5L_0\ /\ (5.l_0)$ from relevant male and female abridged life tables (Chap. 4, Eq. 4.68), and fills in the missing projected numbers of males and females surviving to ages 0–4 at that date. Thus:

$$ {}_5s_0 = B(s)_{t,t+5}\ .\ ({}_5L_0/\,(5\ .\ l_0)) \tag{9.27}$$

Where the equation is applied separately for males and females; ${}_5s_0$ means survivors aged 0–4; $B(s)_{t,t+5}$ stands for births of the relevant sex during the projection interval; life table functions come from the abridged life table assumed to summarize mortality conditions for the relevant sex; ${}_5L_0 = {}_1L_0 + {}_4L_1$.

Having now projected our base-year population five years into the future, we can treat that projected population as a new base population and project ***it*** five years ahead in precisely the same fashion. Thus, in Fig. 9.2 the mid-year population in 'Base year + 5' is treated as the new base population to project a further five years ahead and obtain a projected population for 'Base year + 10'. This process can then be repeated for as many additional five-year projection intervals as we wish (or are game) to add to our overall projection, so that iteratively an n-year projection is generated (where n is some multiple of 5). As with projections over one-year intervals using single-year-of-age data, life tables and fertility schedules used may each either change from projection interval to projection interval to allow for assumed changes in mortality and/or fertility conditions over time, or not change (constant mortality and/or fertility conditions).

Introducing Migration

While the population of the world is closed, major regional populations, national populations and sub-national populations at various geographic levels typically are not, so that projection exercises need also to take account of migration. ***Migration is generally harder to forecast accurately than either mortality or fertility***, especially for small areas. Difficulties in forecasting it derive from a number of things: the greater responsiveness of migration than either fertility or mortality to changing economic conditions, employment prospects, housing availability, transport options and neighbourhood conditions; its susceptibility to influence by government policy, natural disasters and social or cultural conflicts; and overlapping with this its sensitivity, especially at more local levels, to unpredictable events (like sudden closures of major sources of employment) and administrative or legislative actions that introduce or remove what George et al. (2004) call 'special populations' – the likes of refugee groups, university students, military personnel and prison inmates.

For a lot of national populations the net migration component of population change is typically small compared to the birth and death components. At the same time, as birth rates have sunk well below replacement level and life expectancies have risen to new highs in many more developed populations, migration has become a more significant source of population change. It is also ***a less predictable source*** – considerations such as HIV/AIDS aside, the momentum of fertility and mortality trends in most parts of the world over recent decades has been downward and increasingly predictable, whereas migration trends have no such inherent directionality and the migration contribution to population change can fluctuate substantially over comparatively short periods. This lack of predictability may discourage simple extrapolation of past net migration trends and lead to a conservative widening of the range between 'high' and 'low' estimates of future migration in projections.

It is also important to appreciate that ***migration activity tends to be greatest among young adults***. These, of course, are ***people of reproductive age***, so that substantial migration during a projection interval, whether inward or outward in net terms, has important flow-on implications for projected births during both the current and subsequent projection intervals. Get your migration assumptions radically wrong and the natural increase component of your projection is likely to be compromised as well. Failure to predict migration trends acceptably accurately is a common reason for revising population projections sooner than originally planned.

A complication with projecting migration is the reality that while national population projections need to take account only of international migration, sub-national projections (e.g., for the States and Territories of Australia) need to incorporate ***both international and internal migration***. These two types of migration often draw on different data sources and may need to be separately built into sub-national projections.

A range of methodologies exist for incorporating migration into population projections. These variously ***project either net migration flows or gross migration flows*** (i.e., in-migration and out-migration separately). Both approaches are widely used, the latter being theoretically and computationally more elegant, but requiring more data and being more complicated to apply. Methods focused on gross migration flows are often used to deal with internal migration in sub-national population projections. George et al. (2004) describe two basic techniques used in preparing State projections for the United States. The first, employing five-year age groups and projection intervals, uses census data on place of residence five years previously to compute age-sex-specific out-migration ***rates*** for each State (based on estimated populations in age-sex groups five years before the census). These are then applied to base-year populations in age-sex groups to yield a projection of the total pool of interstate out-migrants for all States by age and sex, that pool then being distributed to receiving States according to the ***proportions*** of interstate in-migrants in age-sex groups that each State received during the five-year period prior to the census. Note the distinction between using out-migration ***rates*** and in-migration ***proportions***. It reflects the fact that while members of a population are at risk of migrating ***from*** that population, they are not at risk of migrating ***to*** it. In-migrants are sourced from ***other*** populations.

The second technique for projecting migration based on gross flows uses what are known as multiregional models (Rogers 1985, 1995). In these, migration is treated as part of an integrated system of mortality, fertility and origin-destination population flows between regions of a larger geographic entity (typically a country) by age and sex (and sometimes other demographic attributes as well). Thus, interstate migration in a multiregional model for the United States can be represented by a 51×51 matrix (Washington, D.C. is included) showing, by age and sex, the number of people moving from each State to every other State during the five-year period preceding a census. Migration rates are calculated for each State by dividing destination-specific gross migration flows by age and sex by relevant origin State populations, yielding 50 sets of age-sex-specific out-migration rates, one for each destination State. These rates can then be assumed to be replicated or modified in some prescribed way over subsequent five-year projection intervals.

Projections of net migration use a ***top-down*** or ***bottom-up*** approach, or both. The former first projects total net migration based on recent levels, historical trends, structural models, or some other method, then disaggregates by age and sex based on past observed distributions. The focus is initially broad, then becomes more detailed. A bottom-up approach works the other way round. Net migration rates are developed for each age-sex cohort, then applied to age-sex groups in the base population. The total volume of net migration projected for a population is then obtained as the sum of individual values projected for age-sex groups.

A top-down approach to projecting net migration is often taken in national population projections, where the migration in question is exclusively international migration. Where net migration models are used in making sub-national projections they generally combine international and internal migration. Note that because net migration models do not base migration 'rates' on a population at risk, when used for sub-national projections they are apt to result in inconsistencies in projections for a group of areas (George et al. 2004). Those for areas with rapidly growing populations can be too high, and those for areas with slow or negative growth can be too low. Various constraints can be built into models to prevent this sort of inconsistency getting out of hand.

Given the variety of different approaches to incorporating migration into population projections (see Wilson and Rees (2005) for a review of more recent methodological developments) there is obviously diversity in the precise mathematics employed. Moreover, these days there is rarely a need to carry out projections laboriously by hand, as there are any number of computer packages that will do the job for you provided requisite input data are entered in a prescribed manner. It is therefore not proposed to delve deeply into the mathematics of incorporating migration into a population projection. However, for illustrative purposes we will extend the basic methodology already outlined for projection of a closed population using five-year age groups and five-year projection intervals (Eqs. 9.21, 9.22, 9.23, 9.24, 9.25, 9.26, and 9.27).

It was noted above that while it makes sense to talk of ***rates*** of out-migration from a sub-national population or of emigration from a national population, notions of rates of in-migration or immigration make no sense because the migrants are drawn from outside the focal population. Furthermore, immigration is typically affected by immigration policies which set targets or ceilings in terms of desired or maximum ***numbers*** of migrants, not rates of migration. These sorts of considerations mean that migration assumptions in population projections, especially national population projections where internal migration is not an issue, are usually formulated in terms of absolute numbers rather than rates.

Another key consideration is the reality that migration continuously alters the populations at risk of dying and giving birth throughout a projection interval. In national projections emigrants cease to be at risk of dying and female emigrants cease to be at risk of giving birth as members of the focal population after leaving it; immigrants may pass away between arrival and the end of a projection interval while female immigrants may have children during this period who survive to the end of the projection interval. How do we take this sort of complexity into account?

One convenient approach is to assume that migration in both directions for any sub-category of the population is evenly distributed through the projection interval, which in turn permits an assumption that half moved at the very beginning of the interval and the other half at its very end. The errors introduced (some migrants being deemed to have moved earlier than they actually did and others to have moved later than they did) cancel out.

Suppose, using life table-type notation, we denote net immigration ***of a given sex*** in the five-year age group commencing at exact age x during a five-year projection interval t to t + 5 by ${}_5i_x$. Then to calculate survivors at the end of the projection interval taking account of migration under the assumption outlined above we need to modify Eq. 9.21 by (i) adding directly at the end of the interval half of the net immigration between exact ages x and x + 5 and (ii) adding at the beginning of the interval half of the net immigration between exact ages x-5 and x and surviving that immigrant population over the five-year projection interval. Note that we add immigrants five years younger at the beginning of the projection interval than at the end because the cohort ages five years during the projection interval. Note also that net immigration values may be negative if net ***emigration*** prevails. Our revised version of Eq. 9.21 making allowance for net migration becomes:

$$ {}_5s_x = \left({}_5p_{x-5} + {}^1\!/_2 \cdot {}_5i_{x-5}\right) \cdot \left({}_5L_x/{}_5L_{x-5}\right) + {}^1\!/_2 \cdot {}_5i_x \qquad (9.28) $$

Where ${}_5s_x$ gives survivors at the end of the projection interval in the five-year age group commencing at exact age x (x = 5, 10, 15, etc.) with migration taken account of; ${}_5i_x$ = net immigration during the projection interval between exact ages x and x + 5; other elements have the same meanings as in Eq. 9.21.

Equation 9.28 is applied separately for males and females. In similar fashion Eq. 9.22 for survivors in the open age interval needs modification (and also to be applied separately for each sex).

$$ s_{x+} = \left(p_{(x-5)+} + {}^1\!/_2 \cdot i_{(x-5)+}\right) \cdot \left(T_x/T_{(x-5)}\right) + {}^1\!/_2 \cdot i_{x+} \qquad (9.29) $$

Where x is the exact age marking the lower bound of the open age interval; i_{x+} = net immigration of the relevant sex during the five-year projection interval at ages x and older; T_x is taken from the abridged life table summarizing mortality conditions assumed to prevail for that sex during the projection interval; $p_{(x-5)+}$ = the base-year population of the relevant sex aged x−5 or older.

The number of births during the projection interval also needs adjusting to take account of migration. Female migration increments at the end of the projection interval do not contribute to the number of births during the interval, but those at the beginning of it are assumed to bear children during the interval at the same rate as the population they join.

Recall that Eq. 9.24 gives the number of births during the five-year projection interval as:

$$\mathrm{B}_{\mathrm{t,t+5}} = \Sigma_{\mathrm{x=15-19,45-49}}\,\mathrm{F}\,(\mathrm{x})\ .\ 5\ .\ {}^{1}\!/_{2}\ \ (\mathrm{p(f,x)_t} + (\mathrm{p(f,x-1)_t}\ .\ ({}_5\mathrm{L_x}/{}_5\mathrm{L_{x-5}})))$$

We adjust for migration by substituting for $\mathrm{p(f,x)_t}$ in this equation the quantity $\mathrm{p(f,x)_t} + ½\ \mathrm{i(f,x)}$, where i(f,x) is net female immigration during the five-year projection interval in five-year age group x. Similarly $\mathrm{p(f,x-1)_t}$ becomes $\mathrm{p(f,x-1)_t} + ½\ \mathrm{i(f,x-1)}$, where x−1 means the five-year age group immediately younger than age group x. With these substitutions it transpires that the ***adjustment*** to the number of births to take account of migration (i.e., the quantity to be added to the number yielded by Eq. 9.24) is given by:

$$\Delta\mathrm{B}_{\mathrm{t,t+5}} = \Sigma_{\mathrm{x=15-19,45-49}}\,\mathrm{F}\,(\mathrm{x})\ .\ 5\ .\ {}^{1}\!/_{4}\,(\mathrm{i\,(f,x)} + (\mathrm{i\,(f,x-1)}\ .\ ({}_5\mathrm{L_x}/{}_5\mathrm{L_{x-5}}))) \tag{9.30}$$

This quantity will be negative if net emigration is occurring, reflecting a loss of births due to a loss of potential mothers. The adjusted total number of births is given by:

$$\mathrm{AB}_{\mathrm{t,t+5}} = \mathrm{B}_{\mathrm{t,t+5}} + \Delta\mathrm{B}_{\mathrm{t,t+5}} \tag{9.31}$$

Where $\mathrm{B}_{\mathrm{t,t+5}}$ and $\Delta\mathrm{B}_{\mathrm{t,t+5}}$ are obtained from Eqs. 9.24 and 9.30 respectively.

Adjusted numbers of female births and male births are then obtained from equations modelled on Eqs. 9.25 and 9.26:

$$\mathrm{B(f)}_{\mathrm{t,t+5}} = (1/\,(1 + \mathrm{SRB}))\ .\ \mathrm{AB}_{\mathrm{t,t+5}} \tag{9.32}$$

And

$$\mathrm{B(m)}_{\mathrm{t,t+5}} = (\mathrm{SRB}/\,(1 + \mathrm{SRB}))\ .\ \mathrm{AB}_{\mathrm{t,t+5}} \tag{9.33}$$

Where $\mathrm{B(f)}_{\mathrm{t,t+5}}$ and $\mathrm{B(m)}_{\mathrm{t,t+5}}$ are female and male births, respectively, during the projection interval; SRB stands for 'sex ratio at birth'.

We then apply Eq. 9.27 twice to survive the adjusted number of births of each sex during the five-year projection interval to the end of that interval – i.e., to age group 0–4 years.

What have been developed above are a series of equations for building approximate migration effects into a national population projection using estimates of net immigration. The methodology should be regarded as basic and illustrative. Various alternative approaches (e.g., using separate immigration and emigration flow data) and complications (e.g., different mortality and fertility regimes for migrants compared to non-migrants) are conceivable. It is also possible to recognize other structural variables besides age and sex. The cohort-component approach to population projection can deal with additional structural variables, provided they pertain to attributes acquired at birth (such as ethnic origin or race, although assumptions need to be made in allocating births to ethnic or racial categories given the reality that some will be of mixed parentage). If other structural variables pertaining to characteristics that change during the life course are of interest, multistate methodologies that explicitly recognize age-sex patterns of transitions between states (e.g., labour force status, marital status) and require longitudinal data are likely to be required.

Projection Packages

As previously noted, in the early years of its development cohort-component population projection was a very tedious undertaking. More recently, however, modern computing has spawned a considerable array of computer packages for population projection, most following the cohort-component approach. The following are some examples.

MortPak for Windows (Version 4.3)

The MortPak suite of demographic procedures was developed by the United Nations Population Division (UNPD) as an aid to undertaking demographic analyses for developing country populations. It was first released in 1988 in two versions – MortPak for use on mainframe computers and MortPak-Lite for use on microcomputers (United Nations 1988). These initial versions did not include a projection routine, rather providing 16 programs principally for analysing mortality data, although two were fertility analysis routines and another estimated the completeness of a census compared to a second census. The UNPD had separately made available a population projection program earlier than this (United Nations 1982), but in introducing version 3.0 of MortPak and MortPak-Lite in 1990 added a projection routine to the other 16 (United Nations 1990). MortPak for Windows was launched as version 4.0 in 2003, and the current version, version 4.3, incorporating three further mortality routines, was released in 2013 (United Nations 2013).

The PROJCT routine in version 4.3 carries out a single-year projection by age and sex for up to 100 years into the future based on base-year male and female populations in five-year age groups. These input data are converted to the necessary single-year age group data using an interpolation method known as Beers multipliers (Beers 1944, 1945). Thus five-year age group input data do not necessarily mean that only a projection at five-year intervals can be carried out. The date to which the base population pertains and the final projection year need to be nominated, as do the open age group for the two base populations (minimum is 65+; maximum is 85+) and the sex ratio at birth (minimum 0.75; maximum 1.50; recommended default in the absence of a trustworthy empirical figure 1.05).

Other input data required are, first, male and female life expectancies at birth for at least the base year and the final projection year. Life expectancies at birth may also be given for intermediate projection years, but in their absence the routine interpolates linearly between the base and final projection year values. It also interpolates linearly between successive pairs of life expectancies when intermediate values are given. The age pattern of mortality is provided as a United Nations or Coale-Demeny model life table or a user-designated empirical life table (the UN model life tables accommodate life expectancies at birth up to 92.5 years), and the MortPak routine UNABR is then used to generate single-year probabilities of dying (${}_1q_x$) for every projection year from probabilities for standard abridged life table age groups (exact ages 0–1, 1–5, 5–10, 10–15, etc.). These facilitate the calculation of single-year survival ratios, and hence survivors to single-year age group a, at the end of each one-year projection interval.

Fertility assumptions are built in by providing, at a minimum, total fertility rates and age-specific fertility rates (five-year age groups 15–19 to 45–49) for the base year and the final projection year. TFRs may also be provided for intermediate years at the discretion of the user; values for intermediate years for which data are not provided are calculated by linear interpolation. Age-specific fertility rates for intermediate years are also calculated by linear interpolation with reference to corresponding TFR values.

Net male and female migration needs to be given for the initial and final projection years, and again values may also be given for intermediate years at the user's discretion. Linear interpolation is used to establish values for projection years for which no data are given. Male and female patterns of net migration by age (five-year age groups up to and including the nominated open age group) also need to be provided. These patterns are assumed to apply unchanged throughout the projection period.

PROJCT also requires the user to nominate a print cycle – a value of 1 provides annual projections as output; a value of 5 provides projections for every fifth year beyond the base year. For further information see United Nations (2013: 56–60).

DemProj

DemProj is a population projection program produced by the United States Agency for International Development as part of its *Spectrum* suite of software to inform the establishment and monitoring of development programs and policies (USAID 2008). Other programs in the suite (i) project family planning requirements needed in order to achieve national fertility goals, (ii) compare the costs and benefits of family planning programs, (iii) project the consequences of an AIDS epidemic, (iv) project the social and economic consequences of high fertility and rapid population growth for the labour force, education, health, urbanization and agriculture, (v) estimate the consequences of adolescent reproductive behaviour, and (vi) examine the costs and benefits of interventions to prevent the mother-to-child transmission of HIV.

After choosing a base date and the length of the desired projection period (maximum 150 years) DemProj's first input requirement is the base population by five-year age groups and sex, with a terminal open age group of 80+ years. As with PROJCT in MortPak, these data are converted by DemProj to single-year-of-age data using Beers multipliers (Beers 1944, 1945), facilitating the generation of male and female projections at single-year intervals.

The fertility input required is a total fertility rate and an age distribution of fertility for each year of the projection. TFRs can be input (i) year by individual year, (ii) as a constant value throughout the projection period or defined portions thereof, (iii) by nominating base year and end of projection values and interpolating linearly between them to establish intermediate values, or (iv) by nominating values for selected years during the projection period and interpolating linearly between successive pairs of these to establish intermediate values. Percentage distributions of total fertility can be specified in any of three ways: (i) by selecting one of four United Nations model fertility schedules (known as 'UN sub-Saharan Africa', 'UN Arab', 'UN Asia' and 'Average' (Table 9.1)), (ii) by entering one's own distributions, or (iii) by using the Coale-Trussell fertility model. Choosing a United Nations model is the default option, although as the names of the models suggest they are geared to projecting developing populations (which DemProj was designed for) and are unlikely to be suitable if a more developed population was being projected. The models were originally constructed in 1973 in the context of the UN's regular (nowadays biennial) *World Population Prospects* global projections exercise, and provide age distributions of fertility for TFRs ranging from 7.0 to 2.0 in decrements of 1.0, with distributions for more precise TFRs established by linear interpolation. Stover and Kirmeyer (2007) warn that the selection of a UN model should not be based solely on geography, noting that (p. 21) 'regions are not as well associated with fertility patterns as they are with mortality patterns [because fertility is] subject to relatively more forces that are primarily not biological'. They advise first checking a country's or region's base year age distribution of fertility against all of those given in Table 9.1 to select the one that is most appropriate, rather than choosing blindly on the basis of regional geographic location.

The option in DemProj to enter one's own five-year age group distributions of fertility has the same input options as were listed above for specifying TFRs. Separate distributions can be input for each year; a distribution input for a given year can be kept constant over any subsequent portion of the projection period; or linear interpolation can be used to generate intermediate distributions between pairs of distributions specified for years two or more years apart. This option is especially likely to be attractive if a more developed population is being projected. Table 9.2 shows model age patterns of fertility used by the United Nations Population Division in carrying out its 2012 revision of *World Population Prospects* for two groups of 'low fertility' countries described as the 'market economy countries of Europe' and the 'countries with economies in transition' (United Nations 2014). The former were used for countries such as the Netherlands and also several low fertility countries

Table 9.1 United Nations model tables of the age distribution of fertility

	Age group							
TFR	15–19	20–24	25–29	30–34	35–39	40–44	45–49	Total
Sub-Saharan Africa								
2	8.2	35.4	29.9	17.4	7.2	1.7	0.1	100
3	14.0	31.1	24.7	16.6	9.2	3.6	0.6	100
4	14.9	25.9	22.1	17.1	11.7	6.4	1.8	100
5	16.1	25.4	22.0	17.0	11.6	6.2	1.6	100
6	16.4	24.7	22.1	17.3	11.7	6.2	1.5	100
7	14.7	23.5	21.9	17.9	12.8	7.2	2.0	100
Arab Countries								
2	7.2	31.1	30.3	19.7	9.0	2.4	0.2	100
3	6.6	29.1	29.8	20.7	10.4	3.2	0.2	100
4	7.6	24.4	26.0	21.1	14.2	6.9	1.4	100
5	8.5	23.1	24.9	21.0	14.2	6.9	1.4	100
6	8.8	21.9	24.3	21.1	14.8	7.5	1.6	100
7	7.8	21.7	25.1	21.9	15.0	7.2	1.4	100
Asia								
2	2.8	31.1	38.4	21.1	5.9	0.7	0.0	100
3	2.4	23.5	33.7	25.6	11.9	2.8	0.1	100
4	3.8	20.8	27.9	24.6	15.7	6.3	0.8	100
5	5.6	21.4	26.6	23.3	15.4	6.7	1.0	100
6	7.9	22.8	26.2	22.0	14.2	6.1	0.9	100
7	11.8	24.1	24.1	19.5	13.0	6.3	1.3	100
Average								
2	6.1	32.5	32.9	19.4	7.4	1.6	0.1	100
3	7.7	27.9	29.4	21.0	10.5	3.2	0.3	100
4	8.8	23.7	25.3	20.9	13.9	6.5	1.3	100
5	10.1	23.3	24.5	20.4	13.7	6.6	1.3	100
6	11.0	23.1	24.2	20.1	13.6	6.6	1.3	100
7	11.4	23.1	23.7	19.8	13.6	6.9	1.6	100

Source: Stover and Kirmeyer (2007: 19)

outside Europe (almost certainly including the U.S.A., Canada, Australia and New Zealand) while the latter countries included many of the countries of Eastern Europe (e.g., Slovenia). A separate set of models was used for ‘high fertility’ and ‘medium fertility’ populations (not shown; see United Nations (2014: Table II.2)). Note that the models in Table 9.2 do not differentiate by TFR. All are models applicable to ‘low fertility’ populations, and the skill in using them is to (i) identify the model that best matches known base-year conditions and (ii) also identify any model(s) to which one believes the population in question might trend at some nominated future date(s). Assumptions of constancy or linear change in these selected age patterns of fertility over prescribed periods can then be built into a projection. The ‘market economy countries of Europe’ age patterns of fertility tend, of course, to be much

Table 9.2 Model age patterns of fertility used by the United Nations for projecting low fertility populations in 2012

	Age group								
Model	15–19	20–24	25–29	30–34	35–39	40–44	45–49	Total	Mean age at childbirth
Market economy countries of Europe									
1	2.1	22.9	43.2	26.2	5.2	0.2	0.0	100	28.0
2	1.5	17.5	40.4	31.4	8.7	0.6	0.0	100	29.0
3	1.0	13.2	36.3	35.3	13.0	1.3	0.0	100	30.0
4	0.6	9.8	31.6	37.6	17.9	2.5	0.0	100	31.0
5	0.4	7.2	26.7	38.1	23.0	4.5	0.1	100	32.0
Countries with economies in transition									
1	7.9	35.3	38.4	15.9	2.4	0.1	0.0	100	26.0
2	5.6	29.5	39.3	21.0	4.4	0.2	0.0	100	27.0
3	4.0	24.1	38.4	25.6	7.3	0.6	0.0	100	28.0
4	2.8	19.4	36.2	29.5	10.8	1.3	0.0	100	29.0
5	2.0	15.4	33.1	32.1	14.8	2.5	0.1	100	30.0

Source: United Nations (2014: Tables II.3 and II.4)

older than either those of the 'countries with economies in transition' or those in Table 9.1 (and also the 'high fertility' and 'medium fertility' models used by the UN more recently (United Nations 2014: Table II.2)) – i.e., fertility is less concentrated below age 25 and distinctly more concentrated above age 30.

The Coale-Trussell option for specifying age patterns of fertility uses the ***relational*** attribute of the Coale-Trussell fertility model discussed in Chap. 6 to generate future age patterns of fertility from a base year set of age-specific fertility rates and projected TFRs for following years. It generates patterns that take into account characteristics and components of the base year age-specific rates that reflect deviation from an underlying natural pattern of fertility attributable to the population's marriage pattern and degree of fertility control. According to Stover and Kirmeyer (2007: 26), this option 'is recommended for populations whose initial fertility distributions do not resemble any regional pattern, or have some idiosyncrasy.' It 'performs best in the medium run, if moderate levels of fertility are targeted; or in the short run, regardless of levels of fertility.' However, it is not recommended for projecting low levels of fertility following a substantial period of fertility decline, in which circumstance it is apt to yield distributions that are 'too peaked' (Stover and Kirmeyer 2007: 26).

By default DemProj sets the sex ratio at birth to 105 males per 100 females, but an alternative more precise empirically derived value is also permissible input. Mortality assumptions require the further input of base year and assumed projection period male and female life expectancies at birth, and the nomination of a model life table to generate age patterns of mortality from these. On the former front, once again separate values can be input for each year, or a value input for a given year can be kept constant over any subsequent portion of the projection period, or linear interpolation can be used to generate intermediate values between pairs of life expectancies specified for years two or more years apart. Thus, for example, some sort of linear trend followed by constancy beyond a nominated future date can be accommodated. The model life tables available are the Coale-Demeny West, North, East and South alternatives, the United Nations General, Latin American, Chilean, South Asian or East Asian options, or a customized table supplied by the user.

Migration assumptions require input of the net total number of migrants by sex for each year and the separate specification of male and female age distributions of net migration. Net out-migration is input as a negative number; net in-migration as a positive number. The previously described input options for specifying these assumptions across the projection period are again available. If all age groups for a given sex are assumed to contribute net flows in the same direction as the overall net flow, age distributions of net migration are simple percentage distributions. However, the DemProj manual (USAID 2008) is silent on how age distributions of net migration should be specified if there is reason to assume certain age groups contribute net counterflows to the overall net flow. Presumably such a situation is dealt with via a mixture of positive and negative age-specific percentages, the sum of which is +100. Stover and Kirmeyer (2007: 35) make the point that ‘There are no simple model tables for patterns of migration by age.’ Just as projected net migration estimates themselves need to be based on historical data for the population being projected for a period deemed appropriate leading up to the base year (in conjunction, perhaps, with expressed government policy covering the projection period or at least the earlier part of it), so, too, appropriate sex-specific assumed age patterns of net migration need to be based on recent historical experience of that population.

DemProj also allows for what it calls ‘regional assumptions’. This facility allows for a binary distinction between sub-populations. Perhaps most commonly this is a rural-urban distinction, but the term ‘regional’ can apply to any two-way distinction – e.g., a geographic distinction between highlands and lowlands, a cultural distinction between the native-born and foreign-born populations or indigenous and non-indigenous populations, or a political distinction between the north and south of a country. There are two ways of entering a ‘regional’ assumption. You can ‘directly choose’ the percentage of the entire population you want in one of the two subcategories (e.g., urban) in each year of the projection. Or you can base the sub-population projections on the base year ‘growth rate difference’ between the two sub-populations. This requires input of the base year growth rates for both sub-populations.

For further detail on DemProj see the program manual (USAID 2008). In addition, Stover and Kirmeyer (2007) provide further insight into certain program functionalities and assumptions.

References

Beers, H. S. (1944). Six-term formulas for routine actuarial interpolation. *The Record of the American Institute of Actuaries, 33* Part II (68), 245–260.

Beers, H. S. (1945). Discussion of papers presented in the Record, No. 68: ‘Six-term formulas for routine actuarial interpolation’ by Henry S. Beers. *The Record of the American Institute of Actuaries, 34* Part I (69), 59–60.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Cannan, E. (1895). The probability of the cessation of the growth of population in England and Wales during the next century. *The Economic Journal, 5*, 506–515.

George, M. V., Smith, S. K., Swanson, D. A., & Tayman, J. (2004). Population projections. In D. A. Swanston, J. S. Siegel, & H. S. Shryock (Eds.), *The methods and materials of demography* (pp. 561–601). San Diego: Elsevier Academic Press.

Hinde, A. (1998). *Demographic methods*. London: Arnold.

Keyfitz, N. (1987). The social and political context of population forecasting. In W. Alonso & P. Starr (Eds.), *The politics of numbers* (pp. 235–258). New York: Russell Sage.

Lutz, W. (1996). *The future population of the world: What can we assume today?* London: Earthscan Publication Limited.

McCleary, R., & Hay, R. A. (1980). *Applied time series analysis for the social sciences*. Beverly Hills: Sage.

Notestein, F. W. (1945). Population: The long view. In T. W. Shultz (Ed.), *Food for the World* (pp. 36–69). Chicago: University of Chicago Press.

O’Neill, B. C., Balk, D., Melanie, B., & Ezra, M. (2001). A guide to global population projections. *Demographic Research, 4*(8), 203–288.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modelling population processes*. Oxford: Blackwell.

Rogers, A. (1985). *Regional population projection methods*. Beverly Hills: Sage.

Rogers, A. (1995). *Multiregional demography: Principles, methods and extensions*. Chichester: Wiley.

Rowland, D. T. (2003). *Demographic methods and concepts*. Oxford/New York: Oxford University Press.

Siegel, J. S. (2001). *Applied demography*. San Diego: Academic.

Smith, S. K., Tayman, J., & Swanson, D. A. (2001). *State and local population projections: Methodology and analysis*. New York: Kluwer Academic/Plenum Publishers.

Stover, J. & Kirmeyer, S. (2007). *DemProj version 4: A computer program for making population projections*. Washington, DC: USAID. Available online at http://data.unaids.org/pub/Manual/2007/demproj_2007_en.pdf. Accessed 7 Nov 2014.

United Nations. (1982). *A user’s manual to the population projection computer programme of the population division of the United Nations*. ESA/P/WP.77, New York: United Nations.

United Nations. (1988). *MortPak: The United Nations software package for mortality measurement: Batch-oriented software for the mainframe computer*. ST/ESA/SER.R/78, New York: United Nations. Available online at http://www.un.org/esa/population/publications/MortPak_SoftwarePkg/MortPak_SoftwarePkg.htm. Accessed 25 Oct 2014.

United Nations. (1990). *MortPak and MortPak-Lite upgrades: Version 3.0 of the United Nations software packages for mortality measurement*. ST/ESA/SER.A/117, New York: United Nations.

United Nations. (2013). *MortPak for Windows (Version 4.3)*. POP/SW/MORTPAK/2003, New York: United Nations. Available online at http://www.un.org/en/development/desa/population/publications/pdf/mortality/mortpak_manual.pdf. Accessed 25 Oct 2014.

United Nations. (2014). *World population prospects – The 2012 revision: Methodology of the United Nations population estimates and projections*. New York: United Nations. Available online at http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2012_Methodology.pdf. Accessed 7 Nov 2014.

USAID. (2008). *DemProj: A computer program for making population projections*. Washington, DC: USAID. Available online at http://www.healthpolicyinitiative.com/Publications/Documents/1255_1_DemmanE.pdf. Accessed 16 Aug 2014.

Whelpton, P. (1928). Population of the United States, 1925 to 1975. *American Journal of Sociology, 34*, 253–270.

Whelpton, P. (1936). An empirical method for calculating future population. *Journal of the American Statistical Association, 31*, 457–473.

Willekens, F. J. (1990). Demographic forecasting: State-of-the-art and research needs. In C. Hazeu & G. Frinking (Eds.), *Emerging issues in demographic research* (pp. 9–66). Amsterdam: Elsevier Science.

Wilson, T. (2011). *A review of sub-regional population projection methods*. Queensland Centre for Population Research, University of Queensland. Available online at http://gpem.uq.edu.au/SubRegionalProjectionMethodsReview.pdf. Accessed 20 Jan 2015.

Wilson, T. (2014). New evaluations of simple models for small area population forecasts. *Population, Space and Place*. doi:10.1002/psp.1847.

Wilson, T., & Rees, P. (2005). Recent developments in population projection methodology: A review. *Population, Space and Place, 11*, 337–360.

Index

G.A. Carmichael, *Fundamentals of Demographic Analysis: Concepts, Measures and Methods*, The Springer Series on Demographic Methods and Population Analysis 38, DOI 10.1007/978-3-319-23255-3