AN ANALYSIS OF POPULATION LIFETIME DATA OF SOUTH AUSTRALIA 1841 - 1996

P. I. Leppard

Master of Science (Statistics)

School of Applied Mathematics

The University of Adelaide

December 2002

TABLE OF CONTENTS

ABSTRACT

The average length of life from birth until death in a human population is a single statistic that is often used to characterise the prevailing health status of the population. It is one of many statistics calculated from an analysis that, for each age, combines the number of deaths with the size of the population in which these deaths occur. This analysis is generally known as life table analysis. Life tables have only occasionally been produced specifically for South Australia, although the necessary data has been routinely collected since 1842. In this thesis, the mortality pattern of South Australia over the period of 150 years of European settlement is quantified by using life table analyses and estimates of average length of life.

In Chapter 1, a mathematical derivation is given for the lifetime statistical distribution function that is the basis of life table analysis, and from which the average length of life or current expected life is calculated. This derivation uses mathematical notation that clearly shows the deficiency of current expected life as a measure of the life expectancy of an existing population. Four statistical estimation procedures are defined, and the computationally intensive method of bootstrapping is discussed as an estimation procedure for the standard error of each of the estimates of expected life. A generalisation of this method is given to examine the robustness of the estimate of current expected life.

In Chapter 2, gender and age-specific mortality and population data are presented for twenty five three-year periods; each period encompassing one of the colonial (1841-1901) or post-Federation (1911-96) censuses that have been taken in South Australia. For both genders within a census period, four types of estimate of current expected life, each with a bootstrap standard error, are calculated and compared, and a robustness assessment is made.

In Chapter 3, an alternate measure of life expectancy known as generation expected life is considered. Generation expected life is derived by extracting, from official records arranged in temporal order, the mortality pattern of a notional group of individuals who were born in the same calendar year. Several estimates of generation expected life are calculated using South Australian data, and each estimate is compared to the corresponding estimate of current expected life. Additional estimates of generation expected life calculated using data obtained from the Roll of Honour at the Australian War Memorial quantify the reduction in male generation expected life for 1881-1900 as a consequence of military service during World War I, 1914-18, and the Influenza Pandemic, 1919.

DECLARATION

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying. To assist with this process, a CD-rom containing all data files, computer programs and output (result) files associated with this thesis has been included as an appendix. A printable (post-script) version of the text of the thesis is also contained on the CD-rom.

………………………………..

P. I. Leppard

December 2002

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to thank Dr. C. E. M. Pearce and Dr. G. M. (Mike) Tallis for their guidance in the preparation of this thesis.

The librarians of

- the Barr-Smith Library, The University of Adelaide
- the State Library of South Australia
- the public reading room of State Records of South Australia, especially

were of great assistance in negotiating the historical statistical records of South Australia.

The Registrar of Births, Deaths and Marriages, Ms. Val Edyvean, kindly gave permission to access South Australian death certificates held in the archives of State Records.

Gender and age-specific mortality and population data for South Australian1970-96 was obtained from the Adelaide office of the Australian Bureau of Statistics through Mr. Treva Richards, Information Consultant.

The production of this thesis would not have been possible without the generous co-operation of Dr. K. Baghurst, program manager Consumer Science, Health Sciences and Nutrition, CSIRO, in allowing flexible employment conditions. My colleagues at Consumer Science, Julie Syrette and Sally Record, patiently assisted and guided me through many of the technical details of PC software.

I also wish to gratefully express my appreciation to Mike Tallis for his efforts as teacher, colleague and friend over a period spanning more than thirty years of statistical collaboration.

INTRODUCTION

In 1997 Tallis & Leppard [1] reported the results of a study in a human population, of the predictability of the length of life (lifetime) of a son from the lifetimes of his parents. In this study, a sampling scheme was used in which the records of the South Australian Registry of Births, Deaths and Marriages were randomly accessed to provide a sample of 911 biological families, with an observed lifetime for the son, and an observed lifetime for one or both of the mother and father of each family group. The years of birth ranged from 1874 to 1946 for the sons, with an average lifetime of 68.7 years; from 1834 to 1912 for the mothers, with an average lifetime of 71.5years; and from 1822 to 1916 for the fathers, with an average lifetime of 70.6 years. Thus the sample was a mixture of individuals with a wide range in the calendar year of birth. It has previously been observed, and is now generally acknowledged, that the average lifetimes of males and females in Western populations are increasing with calendar year. Thus for the statistical analysis of the within-family relationships between lifetimes, we decided to standardise the observed lifetime data by using population lifetime distributions specific to gender and calendar year of birth. These distributions are contained within what are generally called population mortality Life Tables, which also include the average lifetime from birth as one of a number of population summary statistics. Unfortunately for our purposes, we found that these life tables have only been routinely and regularly produced for South Australia since 1970. Prior to this year there is a small number of South Australian Life Tables pertinent to the last years of the 19th century and the early years of the 20th century. Many of these life tables have been calculated using a methodology that is now recognised as technically deficient. Although enough population mortality data was either available or could be collected to satisfy the analytic requirements of the within-family lifetimes relationship study, it was apparent that the lifetime characteristics of the evolving South Australian population have not been adequately, comprehensively or systematically documented for the years following the British settlement of South Australia in 1836 until the present time of 1996. The prime objective of this thesis is to provide this information and to investigate the statistical properties of the estimates of average lifetime that are calculated from it.

In Chapter 1, the methodology pertinent to a population mortality life table associated with a specified calendar year is established. The derivation presented in this thesis is based on the concept of a system of statistical lifetime cumulative distribution functions, where a different lifetime distribution function is assumed to characterise each distinct population of individuals born within the discrete calendar years prior to the specified calendar year. This notational framework allows the artificial nature of the derived synthesised lifetime distribution function for the specified calendar year, on which a life table is based, to be clearly seen. The average lifetime, or expected life, is determined from this derived distribution function, and the notation employed here indicates how this summary statistic is most likely to be an under-estimate of the true value of expected life in the prevailing population. The qualifier “current” is added to the terminology for expected life obtained in this manner; as an indication that it is defined by the prevailing or current mortality of the specified calendar year, and as a differentiation from another measure of expected life that is presented in Chapter 3. Estimation procedures are also given in this chapter, and the computer intensive statistical procedure known as bootstrapping is specialised to the derived lifetime distribution function to provide a measure of the effect of sampling variation on the estimate of current expected life. This is an issue that has received very little attention in the literature of population life tables. The bootstrap procedure is generalised so that the robustness of the estimate of current expected life can be examined under a variety of conditions.

In Chapter 2, the procedures developed and discussed in Chapter 1 are applied to an extensive compilation of South Australian data appropriate for the estimation of current expected life over the period of 150 years of European settlement. Much of this data is available on the public record, although it is not always readily available or necessarily tabulated in the most suitable form for analysis. Appropriate statistical techniques are used in these latter circumstances. Many of the tables in this chapter showing gender and age-specific number of deaths have never been previously published. The data for these tables have been obtained by individual inspection of, and extraction from, approximately 18,000 death certificates held in the archives of the South Australian Registry of Births, Deaths and Marriages. The data analysed are available on the accompanying CD-rom that is included as an appendix to this thesis. The naming convention and format of the data files, and the computing environment necessary for their extraction, if required, are described in this chapter. A computer program has been written in the computer language FORTRAN to implement the estimation and bootstrapping procedures that are described in Chapter 1. The usage of the computer program is described in this chapter, and the program source code and executable form are included on the CD-rom. The presentation of data, analyses and results is in reverse chronological order, beginning with the most recent data from 1996-97 and moving backwards through time until 1841. This approach was adopted because the overall quality of data progressively decreases from the quality of current-day data, with earlier years having coarser levels of age tabulation and fewer, if any, official figures for comparison with thesis estimates. Data sets are grouped on the basis of within-group similarities and between-group dissimilarities, and these groupings form the sections of this chapter. Selected extractions from the results of the computer analyses of the data in each section are summarised in a standard tabular form, with the complete output files included on the CD-rom.

In Chapter 3, a methodology is presented in which the data described in Chapter2 are arranged in a manner that allows the lifetime distribution function of a hypothetical population of individuals who are born in a nominated calendar year, that is a “generation”, to be approximated at various subsequent times over the complete lifespan of the “generation”. This formulation is designated a generation lifetime distribution function to distinguish it from the current lifetime distribution function discussed in Chapter 1, and the average lifetime determined from the generation lifetime distribution function is denoted as generation expected life. A FORTRAN computer program has been written to estimate generation expected life for any nominated calendar year from 1841 to 1996 and for each gender, and the bootstrap procedure has been used to provide a standard error of the estimate. The program source code and executable form are included on the CD-rom. Consideration is also given to the influence of any extraordinary events that may have occurred within the lifespan of the “generation” and which is not directly measured by routinely collected data. The “generations” of South Australian males born in 1881-1900 are used for illustration, and the effects of military service during World War I, 1914-18 and the Influenza Pandemic, 1919, on generation expected life for 1881-1900 have been quantified in a number of ways.

It is not the purpose of this thesis to attempt to provide estimates of future mortality through mathematical modelling of, and extrapolation from, current mortality rates. While procedures of this type (*e.g.* Spiegelman [2]) could be applied to the data contained on the CD-rom, analysis in this manner is beyond both the scope and interest of this thesis. The use of current expected life as a predictor of future lifetime for an established population is based on an assumption of stationarity in age-specific mortality rates, and that future rates will not change from the corresponding present rates. Of the two types of estimator of expected life presented in this thesis, generation expected life is the closest conceptually to the expected value of the lifetimes of an actual population of individuals. Since it is only possible to calculate generation expected life retrospectively, it therefore cannot be used as a predictor of future lifetime. However, by calculating current and generation expected life for the same calendar year, an examination can be made of the extent by which current expected life mis-estimates future lifetime, as measured by generation expected life. Several comparisons of this kind are given in Chapter 3.

CHAPTER 1: THE RATIONALE AND THEORY OF CURRENT LIFE TABLE ANALYSIS AND CURRENT EXPECTED LIFE

1.1 Introduction

The direct estimation of the statistical distribution of the lengths of life experienced by a human population is not possible by means of a conventional observational study, for two very obvious reasons. Consider a hypothetical prospective cohort study, in which a large group of concurrently born infants is observed over their lifetimes until all have died. Firstly, with unrestricted migration locally, nationally and internationally, there are the consequent problems in monitoring a group of free-living individuals in order to determine the age at which each death occurs. Secondly, there is the fact of longevity of humans. The results of modern censuses indicate that in Western industrialized societies there is an increasing percentage of the general population aged 90 years or more, with some individuals living to extreme old age. For example, it was reported in March 2001 that South Australia’s oldest person had just celebrated her 113th birthday [3], in March 2002 it was reported that the world’s oldest person had died at the age of 115 years [4], and in May 2002 that Australia’s oldest person had died at the age of 114 years [5]. Consequently, two or more generations of investigators would be required before all data became available for analysis following the death of the oldest individual, approximately a hundred years after the commencement of the study. Moreover, any lifetime distribution derived from such a prospective study would only be applicable to a population born a hundred or more years in the past.

Alternatively, a form of retrospective cohort study could be conducted by taking a sample of individuals, selected from official birth records and all born in the same calendar year. Again the longevity characteristic of humans would require that the chosen calendar year be a notional hundred years prior to the year of the study. The lengths of life of these sampled individuals would be found by successive yearly audits of official death records following the year of their birth. The amount of time needed to do this type of study would therefore be, for all intents and purposes, the time expended on the search through the death records, and hence would be feasible. However, the success in identification of all deaths would be dictated by the nature and extent of the system of death records, and failure to identify some or indeed many deaths would introduce statistical bias into the analysis of length of life. The potential extent of this problem can be illustrated by the lifetime relationship study of Tallis & Leppard [1] discussed in the Introduction. The 1,822 parents, with all father-mother pairs assumed to be alive at the time of the birth of their son, had a group average age of 32.6 years. However, a certificate of death was not found for 348 (19%) of the parents following a search of the Registry death records forward in time from the date of birth of the pertinent son. These estimates suggest that by attempting to determine the length of time from birth until death of individuals through a registry system, but for approximately thirty years more than in the Tallis & Leppard study, would result in at least 20% of the group having an unknown length of life. The relevance of any derived lifetime distribution, most probably biased, would also be to a population born a hundred or more years previously.

To overcome these difficulties, a numerical method has been developed to approximate a population lifetime distribution by using the recorded number of deaths and the size of the population in which these deaths occurred. This method is known as life table analysis, and it is generally accepted to have originated independently through the work of John Graunt in 1662 [6] and Edmund Halley in 1693 [7]. Since the method was developed initially for actuarial and life insurance purposes, it is sometimes called actuarial life table analysis. However, to clearly distinguish this specific method from a related analysis that is discussed briefly in the next paragraph, modern terminology is commonly current life table analysis. The qualifier "current" implies mortality data obtained from a particular year that is combined through calculation to represent the lifetime distribution of the population of interest. A derivation of the methodology and a discussion of the interpretation of the results of a current life table analysis, when applied to the estimation of the lifetime distribution of a human population, are presented in the following sections of this chapter.

In contrast, there are other situations where time is measured between two defined events, which may be birth and death but not necessarily so, but which do not have the problems discussed above. For example, in a scientific experiment, each animal of a group may be given a treatment, and the attribute of interest for an animal is the elapsed time until the occurrence of a defined event. The controlled nature of the experiment means that after a predetermined period specified by the experimental protocol, each animal has either a measured time to the event, or is known to have lasted (or "survived") the experiment free of treatment effect. Analyses of this type of data, being a mixture of exact observed survival times and survival times censored to have a value equal to the length of time under study, are collectively called survival analyses. There is a large body of literature and many standard texts in this area; *e.g.* Elandt-Johnson & Johnson [8], and Cox & Oakes [9]. A basic form of survival analysis is called cohort life table analysis. While superficially similar to current life table analysis in both likeness of name and presentation of results, there is a fundamental difference between the two methods. A cohort life table analysis summarizes the actual survival experience of the individuals on which the analysis is based, whereas a current life table analysis expresses cross-sectional mortality experience, derived from a large group of individuals of many different ages, as the lifetime distribution of a hypothetical population. It is this latter form of life table analysis that is the topic of this thesis.

1.2 Current life table analysis

Current life table analysis began in the 17th century for the specific purpose of the determination of life insurance, and the technique has been since refined within the actuarial sciences. However, usage of current life table analysis has spread to the disciplines of demography, health science and vital statistics because it is the most effective means of summarizing the mortality experience of a human population, enabling national trends to be identified and international comparisons made. The most commonly quoted summary quantity arising from a current life table analysis is the expected length of life at birth, which is a convenient single value that is used for comparison purposes amongst groups within and across countries.

There is a large and diverse literature for current life table analysis, using various means to present and justify the methodology. This seemingly reflects the presumed level of mathematical abilities of the general readership in the different areas of application. For example, Spiegelman [2] and Benjamin & Haycocks [10] develop current life table analysis for actuaries using a detailed probabilistic approach. In contrast, Selvin [11] and Newell [12], writing for epidemiologists and demographers respectively, give a less detailed mathematical justification of the underlying principles and concentrate more on the arithmetical operations involved in the analysis. Others like Namboodiri & Suchindran [13], Chiang [14], and Elandt-Johnson & Johnson [8], who are writing for a more advanced mathematical and statistical readership, develop their presentation using probability and integral calculus.

Whatever the method of derivation, the results of a current life table analysis are universally presented as a table of numbers called, not surprisingly, a current or mortality life table. The rows of this table specify whole years of life after birth. A table with an age range from 0 to 85 years with increments of one year is conventionally called a complete current life table. Sometimes ages may be displayed in ranges, typically ranges of 5 years after the age of five, and the resulting table is called an abridged current life table. There are two columns of principal interest in both forms of the current life table. One column indicates the number of individuals alive at each age from a specified initial population size, usually 100,000. The other column gives the conditional expected length of life given the attainment of each age. The notation l_x and e_x, where x indicates completed age in years, is commonly used in published current life tables to refer to these two columns. Thus, in particular, the expected length of life at birth is e_0. Traditionally, tables are calculated separately for males and females because of historically different mortality patterns.

1.3 Derivation of current life table analysis

In this thesis, current life table analysis is established from a statistical perspective by using a system of statistical cumulative distribution functions (*cdf*s). This structure provides a framework that allows an interpretation of the derived results that is otherwise not possible. Conceptually it is assumed that for a population of individuals born at the same calendar time, the realized lengths of time from birth until death (*i.e.* the random variable “length of life”, hereafter X) are described by a *cdf*, denoted generically by the notation $F = F(x) = \Pr(X \leq x)$. Since every length of life is a positive and finite quantity, then values x of X are such that $0 \leq x \leq L$ for some value L. Observational and anecdotal evidence [15] suggest that a reasonable minimal value of L is perhaps of the order of 120 years or so, but irrespectively $F(0) = 0$ and $F(L) = 1$. Since a specific analytic form is not required for F, a current life table analysis can therefore be considered a nonparametric procedure. It is also assumed that F is continuous and has a probability density function (*pdf*) f(x). *i.e.* $F(x) = \int_0^x f(y)\,dy$.

With these specifications, assumptions and notation, three general results are derived for F in Section 1.3.1 to Section 1.3.3: the first two indicate possible ways by which F might be estimated, and the third establishes the relationship between the expected value of X and F.

1.3.1 F expressed as a product of conditional probabilities

Let the partition $P = \{x_0, x_1, ..., x_n \mid x_0 \le x_1 \le ... \le x_n\}$ with additionally $x_0 \ge 0$ and $x_n \le L$.

Then for i =1,2,…,n

$F(x_i)$

$= 1 - [1 - F(x_i)]$

$$= 1 - [1- F(x_{i-1})] \frac{1 - F(x_i)}{1 - F(x_{i-1})}$$

$$= 1 - [1- F(x_{i-1})] \frac{1 - F(x_{i-1}) - [F(x_i) - F(x_{i-1})]}{1 - F(x_{i-1})}$$

$$= 1 - [1- F(x_{i-1})] [1 - \frac{F(x_i) - F(x_{i-1})}{1 - F(x_{i-1})}]$$

Applying operations to $[1-F(x_{i-1})]$ similar to those applied to $[1-F(x_i)]$ and recursively reducing $[1-F(x_{i-2})]$, $[1-F(x_{i-3})]$ *etc* in turn

$$= 1 - [1- F(x_0)][1 - \frac{F(x_1) - F(x_0)}{1 - F(x_0)}][1 - \frac{F(x_2) - F(x_1)}{1 - F(x_1)}]......[1 - \frac{F(x_i) - F(x_{i-1})}{1 - F(x_{i-1})}]$$

1.3.2 F specified by a hazard function

The hazard function u(x) of a random variable X is defined as

$u(x)$

$= f(x) / [1-F(x)]$

$= - d/dx \ln[1-F(x)]$

Then $\int_0^x u(t)\, dt = - \ln [1-F(x)]$, since F(0) = 0, and hence algebraically

$F(x)$

$$= 1 - \exp(-\int_0^x u(t)\, dt)$$

$$= 1 - \exp(-\sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} u(t)\, dt) \qquad \text{properties of integral calculus}$$

$$\approx 1 - \exp(-\sum_{i=1}^{n} (x_i - x_{i-1})\, u(\tfrac{1}{2}(x_{i-1} + x_i))) \qquad \text{Section 1.4.2}$$

for the partition P with additionally $x_0 = 0$ and $x_n = x \le L$.

1.3.3 The expected value of X determined from F

E[X]

$= \int_0^L x\, f(x)\, dx$

$= L\, F(L) \;-\; \int_0^L F(x)\, dx$ integration by parts

$= \int_0^L [1 - F(x)]\, dx$ since F(L) = 1

$= \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} [1 - F(x)]\, dx$

for the partition P with additionally $x_0 = 0$ and $x_n = L$.

This is a special case of a general expression for all moments of a positively valued random variable, stated and proved by Feller [16].

1.4 Mathematical approximations

Simplification of some of the mathematical expressions established in Section 1.5 requires the following three approximation formulae for a function g(x) continuous in the closed interval [a,b]. Details can be found in standard texts: *e.g.* Abramowitz & Stegun [17] and Spiegel [18]

1.4.1 Trapezoidal rule for numerical integration

$$\int_a^b g(x)\, dx \;\approx\; \tfrac{1}{2}(b-a)\,(g(a) + g(b))$$

1.4.2 Mean value rule for numerical integration

$$\int_a^b g(x)\, dx \;\approx\; (b-a)\, g(\tfrac{1}{2}(a+b))$$

1.4.3 Linear interpolation

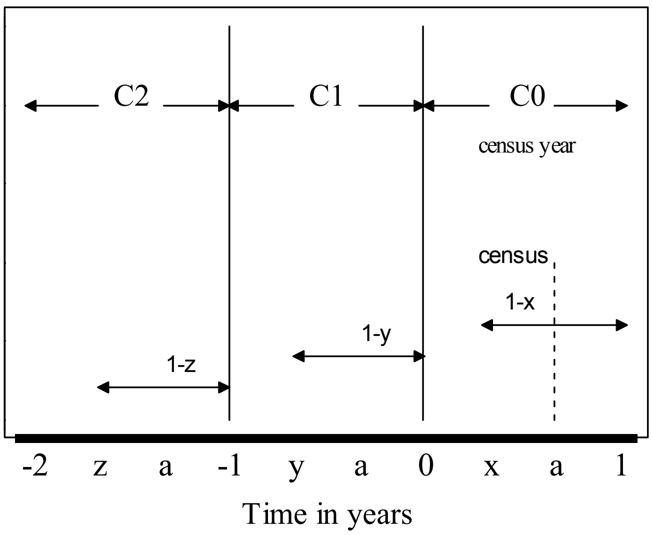
$g(x) \;\approx\; g(a) + (x - a)\,[g(b) - g(a)] / (b - a)$, for $a \le x \le b$ and $g(a) \le g(b)$.

$g(x) \;\approx\; (1-\delta)\, g(a) + \delta\, g(a+1)$, when $b = a + 1$, $x = a + \delta$ and $0 \le \delta \le 1$.

1.5 Age-specific expected number of deaths and expected size of the population

In this section, a relationship is established between F and the age-specific expected number of deaths for a calendar year, and between F and the age-specific expected number of individuals alive at a given time in that same calendar year. Observational data is used in Section 1.7 to estimate both the expected number of deaths and expected population sizes. For this reason, the arguments developed in this section are centred on a notional calendar year in which a population census is taken, and where accordingly estimates of population sizes are readily available. The choice of a time unit of one calendar year is a consequence of circumstances, in that a year is generally the smallest and most frequently used unit for tabulation of death and population figures in official publications.

Figure 1.5: Timescale



With reference to the timescale shown in Figure 1.5, let C_0 indicate a calendar year in which a census is taken at some time “a” within that year, $0 \le a \le 1$. The origin of the timescale is the beginning of C_0, and the negative values shown indicate the beginning of calendar years preceding C_0. For each of these calendar years, the same relative time-point "a" as in C_0 is also shown. Let $N_0(x)$ be the number of births at any time x within C_0, $0 \le x \le 1$, and let F_0 be the particularization of the *cdf* F defined in Section 1.3 that is assumed to be applicable to all individuals born in C_0. The density and hazard functions associated with F_0 are f_0 and u_0 respectively. Similarly, let C_1 be the calendar year immediately prior to C_0, with analogous definitions of $N_1(y)$, $0 \le y \le 1$, F_1, f_1 and u_1; and let C_2 be the calendar year two years previous to C_0, again with analogous definitions of $N_2(z), 0 \le z \le 1$, F_2, f_2 and u_2; and so on for calendar years $C_3, C_4, \ldots$ *etc* (not shown). The subscript notation thus indicates yearly intervals prior to the pivotal census year C_0 and avoids the inconvenience of negative indices in the following mathematical arguments. The populations within each calendar year are considered closed to migration.

This formulation in terms of a sequence of discrete time intervals is of course not meant to imply that any actual underlying lifetime *cdf*s F_0, F_1, F_2 *etc* abruptly change in a step-wise manner at the boundaries of each calendar year. It is a compromise between the possibility of the function F continuously changing over time and mathematical tractability, and allows the effect of any change in F with calendar year to be accommodated with relative ease. Various relationships can be established with these definitions.

Firstly let ${}_1D_0$ denote the expected number of deaths occurring in C_0 of individuals aged between birth and one year. Then it follows that

${}_1D_0$ is

the expected number of deaths occurring in C_0 of individuals aged between birth and one year and who were born in C_0

plus

the expected number of deaths occurring in C_0 of individuals aged between birth and one year and who were born in C_1

$= \int_0^1 N_0(x)\, F_0(1\text{-}x)\, dx + \int_0^1 N_1(y)\,[F_1(1)\text{-}F_1(1\text{-}y)]\, dy$ by reference to Figure 1.5

$\approx \tfrac{1}{2}\, N_0\, F_0(1) + \tfrac{1}{2}\, N_1\, F_1(1)$ assuming $N_0(x) = N_0$ and $N_1(y) = N_1$ for $0 \le x, y \le 1$ and applying Section 1.4.1

$= N_0\, F_0(1)$ [OPTION 1(a)] assuming $N_1 = N_0$ and $F_1 = F_0$

$= N_0 \int_0^1 f_0(t)\, dt$

$\approx N_0\, f_0(.5)$ [OPTION 2(a)] applying Section 1.4.2

Also, let ${}_1P_0$ denote the expected number of individuals alive in C_0 at the time of the census and aged less than one year. Then it follows that

${}_1P_0$ is

the expected number born in C_0 in the time period (0,a) and living to at least time a in C_0

plus

the expected number born in C_1 in the time period (a,1) and living to at least time a in C_0

$= \int_0^a N_0(x)\,[1\text{-}F_0(a\text{-}x)]\, dx + \int_a^1 N_1(y)[1\text{-}F_1(1\text{+}a\text{-}y)]\, dy$ by reference to Figure 1.5

$\approx aN_0 - \tfrac{1}{2} aN_0[F_0(0) + F_0(a)] + (1\text{-}a)N_1 - \tfrac{1}{2}(1\text{-}a)N_1[F_1(a) + F_1(1)]$

assuming $N_0(x) = N_0$, $N_1(y) = N_1$ for $0 \le x, y \le 1$ and applying Section 1.4.1

$= N_0 - \tfrac{1}{2}\, N_0\,[(1\text{-}a)\,F_0(a) + (1\text{-}a)\,F_0(1) + a\,F_0(a)$ assuming $N_1 = N_0$, $F_1 = F_0$ with $F_0(0) = 0$

$= N_0\,\{1 - \tfrac{1}{2}\,[\,F_0(a) + (1\text{-}a)\; F_0(1)]\}$

$\approx N_0 \{1 - \tfrac{1}{2} [F_0(0) + a[F_0(1) - F_0(0)] + (1\text{-}a) F_0(1)]\}$ applying Section 1.4.3 to $F_0(a)$

$= N_0 [1 - \tfrac{1}{2} F_0(1)]$

$= N_0 [1 - F_0(0)] - \tfrac{1}{2} N_0 F_0(1)$ [OPTION 1(b)] since $F_0(0) = 0$

$\approx N_0 [1 - F_0(.5)]$ [OPTION 2(b)] applying Section 1.4.3

With these results, there are now two alternative ways in which to proceed:

Combining the results labelled OPTION 1(a) and OPTION 1(b) above, ${}_1q_0$ is defined as

$${}_1q_0$$

$$= {}_1D_0 \ / \ ({}_1P_0 + \tfrac{1}{2}\, {}_1D_0)$$

$$\approx \frac{N_0 F_0(1)}{N_0[1 - F_0(0)] - \tfrac{1}{2} N_0 F_0(1) + \tfrac{1}{2} N_0 F_0(1)}$$

$$= \frac{F_0(1) - F_0(0)}{1 - F_0(0)}$$

Combining the results labelled OPTION 2(a) and OPTION 2(b) above, $u_0(.5)$ is defined as

$$u_0(.5)$$

$$= {}_1D_0 \ / \ {}_1P_0$$

$$\approx \frac{N_0 f_0(.5)}{N_0[1 - F_0(.5)]}$$

$$= \frac{f_0(.5)}{1 - F_0(.5)}$$

In a similar manner, the expected number of deaths and expected population size of individuals aged between one and two years can be established. Let ${}_1D_1$ denote the expected number of deaths occurring in C_0 of individuals aged between the ages of one and two years. Then it follows that

${}_1D_1$ is

the expected number of deaths occurring in C_0 of individuals aged between the ages of one and two years who were born in C_1

plus

the expected number of deaths occurring in C_0 of individuals aged between the ages of one and two years who were born in C_2

$$= \int_0^1 N_1(y)[F_1(2\text{-}y) - F_1(1)]\, dy \; + \; \int_0^1 N_2(z)[F_2(2) - F_2(2\text{-}z)]\, dz \qquad \text{by reference to Figure 1.5}$$

$$\approx N_1\{\tfrac{1}{2}[F_1(1) + F_1(2)] - F_1(1)\} \quad + \quad N_2\{F_2(2) - \tfrac{1}{2}[F_2(1) + F_2(2)]\}$$

assuming $N_1(y) = N_1$ and $N_2(z) = N_2$ for $0 \leq y\,, z \leq 1$ and applying Section 1.4.1

$$= \tfrac{1}{2} N_1[F_1(2) - F_1(1)] + \tfrac{1}{2} N_2[F_2(2) - F_2(1)]$$

$$= N_1\,[\,F_1(2) - F_1(1) \qquad \text{[OPTION 1(c)]} \qquad \text{assuming } N_2 = N_1 \text{ and } F_2 = F_1$$

$$= N_1 \int_1^2 f_1(t)\, dt$$

$$\approx N_1\; f_1(1.5) \qquad \text{[OPTION 2(c)]} \qquad \text{applying Section 1.4.2}$$

Similarly, let ${}_1P_1$ denote the expected number of individuals alive at the time of the census in C_0 and aged between the ages of one and two years. Then it follows that

${}_1P_1$ is

the expected number born in C_1 in the time period (0,a) and living to at least time a in C_0

plus

the expected number born in C_2 in the time period (a,1) and living to at least time a in C_0

$$= \int_0^a N_1(y)[1 - F_1(1+a-y)]\,dy + \int_a^1 N_2(z)[1 - F_2(2+a-z)]\,dz$$ by reference to Figure 1.5

$$\approx N_1\{a - \tfrac{1}{2}a[F_1(1) + F_1(1+a)]\} + N_2\{(1-a) - \tfrac{1}{2}(1-a)[F_2(1+a) + F_2(2)]\}$$

assuming $N_1(y) = N_1$ and $N_2(z) = N_2$ for $0 \le y\,,\, z \le 1$

and applying Section 1.4.1

$$= N_1 - \tfrac{1}{2}N_1\{(1-a)[F_1(1+a) + F_1(2)] + a[F_1(1) + F_1(1+a)]\}$$ assuming $N_2 = N_1$ and $F_2 = F_1$

$$= N_1 - \tfrac{1}{2}N_1[F_1(1+a) + (1-a)F_1(2) + aF_1(1)]$$

$$\approx N_1\{1 - \tfrac{1}{2}[F_1(1) + F_1(2)]\}$$ applying Section 1.4.3

$$= N_1[1 - F_1(1)] - \tfrac{1}{2}N_1[F_1(2) - F_1(1)]$$ [OPTION 1(d)]

$$\approx N_1[1 - F_1(1.5)]$$ [OPTION 2(d)] applying Section 1.4.3

Combining the results labelled OPTION 1(c) and OPTION 1(d) above, ${}_1q_1$ is defined as

${}_1q_1$

$= {}_1D_1 \,/\, ({}_1P_1 + \tfrac{1}{2}\,{}_1D_1)$

$\approx \dfrac{N_1[F_1(2) - F_1(1)]}{N_1[1 - F_1(1)] - \tfrac{1}{2}N_1[F_1(2) - F_1(1)] + \tfrac{1}{2}N_1[F_1(2) - F_1(1)]}$

$= \dfrac{F_1(2) - F_1(1)}{1 - F_1(1)}$

Combining the results labelled OPTION 2(c) and OPTION 2(d) above, $u_1(1.5)$ is defined as

$u_1(1.5)$

$= {}_1D_1 \,/\, {}_1P_1$

$\approx \dfrac{N_1 f_1(1.5)}{N_1[1 - F_1(1.5)]}$

$= \dfrac{f_1(1.5)}{1 - F_1(1.5)}$

Thus, by repeated application of the above reasoning to calendar years C_3, C_4, C_5 *etc*, increasingly more remote in time prior to C_0, the quantities

${}_1D_k$ = the expected number dying between ages k and k+1 years in C_0

${}_1P_k$ = the expected number alive between ages k and k+1 years at any time a in C_0

are used to define, for k=0,1,2,…

$${}_1q_k = {}_1D_k \,/\, ({}_1P_k + \tfrac{1}{2}\,{}_1D_k) \approx \frac{F_k(k+1) - F_k(k)}{1 - F_k(k)}$$

$$u_k(k+.5) = {}_1D_k \,/\, {}_1P_k \approx \frac{f_k(k+.5)}{1 - F_k(k+.5)}$$

These results form the foundation for the construction of *cdf*s, from which expected values of length of life at birth can be calculated. *(see* Section 1.7)

The relationships established thus far can be generalized to age ranges of more than one year. For example, let ${}_5D_5$ denote the expected number of deaths occurring in C_0 of individuals aged between the ages of five and ten years. Then it follows that

$${}_5D_5$$

$$= {}_1D_5 + {}_1D_6 + {}_1D_7 + {}_1D_8 + {}_1D_9$$

$$\approx N_5 [F_5(6) - F_5(5)] + N_6 [F_6(7) - F_6(6)] + N_7 [F_7(8) - F_7(7)] + N_8 [F_8(9) - F_8(8)]$$

$$+ N_9 [F_9(10) - F_9(9)]$$ using OPTION 1(c) for each term

$$= N_5 [F_5(10) - F_5(5)]$$ [OPTION 1(e)] assuming $N_9 = N_8 = N_7 = N_6 = N_5$ and $F_9 = F_8 = F_7 = F_6 = F_5$

$$= N_5 \int_5^{10} f_5(t)\, dt$$

$$\approx 5\, N_5\, f_5(7.5)$$ [OPTION 2(e)] applying Section 1.4.2

Similarly, let ${}_5P_5$ denote the expected number of individuals alive at the time of the census in C_0 and aged between the ages of five and ten years. Then it follows that

$${}_5P_5$$

$$= {}_1P_5 + {}_1P_6 + {}_1P_7 + {}_1P_8 + {}_1P_9$$

$$\approx N_5 [1 - F_5(5)] - \tfrac{1}{2} N_5 [F_5(6) - F_5(5)] + N_6 [1 - F_6(6)] - \tfrac{1}{2} N_6 [F_6(7) - F_6(6)]$$

$$+ N_7 [1 - F_7(7)] - \tfrac{1}{2} N_7 [F_7(8) - F_7(7)] + N_8 [1 - F_8(8)] - \tfrac{1}{2} N_8 [F_8(9) - F_8(8)]$$

$$+ N_9 [1 - F_9(9)] - \tfrac{1}{2} N_9 [F_9(10) - F_9(9)]$$

using OPTION 1(d) for each term

$= N_5 [5 - F_5(5) - F_5(6) - F_5(7) - F_5(8) - F_5(9)] - \frac{1}{2} N_5 [F_5(10) - F_5(5)]$

assuming $N_9 = N_8 = N_7 = N_6 = N_5$

and $F_9 = F_8 = F_7 = F_6 = F_5$

$\approx N_5 [5 - F_5(5) - \{F_5(5) + \frac{1}{5}[F_5(10) - F_5(5)]\} - \{F_5(5) + \frac{2}{5}[F_5(10) - F_5(5)]\}$

$- \{F_5(5) + \frac{3}{5}[F_5(10) - F_5(5)]\} - \{F_5(5) + \frac{4}{5}[F_5(10) - F_5(5)]\}] - \frac{1}{2} N_5 [F_5(10) - F_5(5)]$

applying Section 1.4.3 for each term $F_5(k)$, k = 6 - 9

$= 5 N_5 [1 - F_5(5)] - \frac{5}{2} N_5 [F_5(10) - F_5(5)]$ [OPTION 1(f)]

$= 5 N_5 \{1 - F_5(5) - \frac{1}{2}[F_5(10) - F_5(5)]\}$

$\approx 5 N_5 [1 - F_5(7.5)]$ [OPTION 2(f)] applying Section 1.4.3

Combining the results labelled OPTION 1(e) and OPTION 1(f) above, ${}_5q_5$ is defined as

${}_5q_5$

$= 5\,{}_5D_5 \,/\, ({}_5P_5 + \frac{5}{2}\,{}_5D_5)$

$$\approx \frac{5N_5[F_5(10) - F_5(5)]}{5N_5[1 - F_5(5)] - \frac{5}{2}N_5[F_5(10) - F_5(5)] + \frac{5}{2}N_5[F_5(10) - F_5(5)]}$$

$$= \frac{F_5(10) - F_5(5)}{1 - F_5(5)}$$

Combining the results labelled OPTION 2(e) and OPTION 2(f) above, $u_5(7.5)$ is defined as

$u_5(7.5)$

$= {}_5D_5 \,/\, {}_5P_5$

$$\approx \frac{5N_5 f_5(7.5)}{5N_5[1 - F_5(7.5)]}$$

$$= \frac{f_5(7.5)}{1 - F_5(7.5)}$$

Thus, by repeated argument, the quantities

${}_jD_k =$ the expected number dying between ages k and k+j years in C_0

${}_jP_k =$ the expected number alive between ages k and k+j years at any time a in C_0

are used to define, for k = 0,1,2,…in combination with conventionally either j = 1 or j = 5

$${}_jq_k = j \; {}_jD_k \,/\, ({}_jP_k + \tfrac{j}{2} \; {}_jD_k) \approx \frac{F_k(k+j) - F_k(k)}{1 - F_k(k)}$$

$$u_k(k + j/2) = {}_jD_k \,/\, {}_jP_k \approx \frac{f_k(k + j/2)}{1 - F_k(k + j/2)}$$

1.6 Some comments about the fundamental estimator $_jq_k$

Quantities of the form, $_jq_k = j\ _jD_k\ /\ (\ _jP_k + \frac{1}{2}\ _jD_k\ )$,with appropriate values of j and k, are fundamental to the construction of a conventional current life table. This definition is invariably obtained irrespective of what mathematical method of derivation is used to justify the result. In some presentations $_jq_k$ has a more general form, with an additional term $_jf_k$, so that $_jq_k = j\ _jD_k\ /\ (\ _jP_k + j(1-{_jf_k})\ _jD_k\ )$; $_jf_k$ being described in Elandt-Johnson [8] as “the expected fraction of the [k,k+j) interval for those aged k who die in [k,k+j)”. Empirical estimates of $_jf_k$ are needed if this formulation is used because values of $_jf_k$ are assumed to be population specific, in the opinion of Chiang [14], or, again from Elandt-Johnson, “assuming uniform distribution of time at death in [k,k+j) , we obtain $_jf_k = \frac{1}{2}$”. In this thesis, with a presentation developed by using a sequence of *cdf*s, the justification of “ $_jf_k = \frac{1}{2}$ ” arises from the mathematical approximations to integrals and other functional expressions described in Section 1.5. The level of accuracy in the approximations to the true values ultimately depends on the mathematical properties of the functions being evaluated and which in this application are unknown. Other types of approximation may be argued that would result in “ $_jf_k \neq \frac{1}{2}$ ”, but these are not considered here.

The expected population size $_jP_k$ is often referred to, for example Newell [12], as “the population aged k at mid-year”. Cox [19] says “As censuses are not normally held on 30 June, however, the enumerated population would probably need some adjustment to convert it to the size of the estimated population….on average over the whole of this period”. In this thesis $_jP_k$ has been defined as the expected size of the population aged between k and k+j years at a designated census time within a nominated calendar year. The time of the census is arbitrary and is not necessarily the mid-point of that year. This apparent difference in definition of $_jP_k$ is reconciled by noting the consequence of the two simplifying assumptions used in this thesis, of constant birth rate and *cdf* over adjacent calendar years, to derive the form of $_jq_k$.

1.7 Definition of current life table analyses using the quantities $_jq_k$

For this thesis, a complete current life table analysis is the combination of the quantities $_1q_0, _1q_1, _1q_2, ..., _1q_{104}$ in the manner suggested by the result of Section 1.3.1 to define a synthesized distribution function G_C for which

$$G_C(k+1) = 1 - \prod_{i=0}^{k}(1 - {_1q_i}) \text{, for } k = 0,(1),104$$

Similarly, for this thesis, an abridged current life table analysis is the combination the quantities $_1q_0, {_1q_1}, {_1q_2}, {_1q_3}, {_1q_4}, {_5q_5}, {_5q_{10}}, ..., {_5q_{100}}$ in the manner suggested by the result of Section 1.3.1 to define a synthesized distribution function G_A for which

$$G_A(k+1) = G_C(k+1) \text{, for } k = 0,(1),4$$

and

$$G_A(k+5) = 1 - \prod_{i=0}^{4}(1 - {_1q_i}) \prod_{i=5,(5)}^{k}(1 - {_5q_i}) \text{, for } k = 5,(5),100$$

The relationship to the conventional notation introduced in Section 1.2 for a complete current life table, for the number of individuals surviving to age x from an initial population size of 100,000 is: $l_x = 100000\,(1 - G_C(x))$, for x = 1,(1),100

It is informative to examine the first three terms of G_C; $G_C(1)$, $G_C(2)$ and $G_C(3)$.

$G_C(1)$

$= 1 - (1 - {}_1q_0)$

$\approx 1 - (1 - \frac{F_0(1) - F_0(0)}{1 - F_0(0)})$

$= F_0(1)$

$G_C(2)$

$= 1 - (1 - {}_1q_0)(1 - {}_1q_1)$

$\approx 1 - (1 - \frac{F_0(1) - F_0(0)}{1 - F_0(0)})(1 - \frac{F_1(2) - F_1(1)}{1 - F_1(1)})$

$= 1 - (1 - F_0(1))\frac{1 - F_1(2)}{1 - F_1(1)}$

$\neq F_0(2)$ unless $F_1(1) = F_0(1)$ and $F_1(2) = F_0(2)$

$G_C(3)$

$= 1 - (1 - {}_1q_0)(1 - {}_1q_1)(1 - {}_1q_2)$

$\approx 1 - (1 - \frac{F_0(1) - F_0(0)}{1 - F_0(0)})(1 - \frac{F_1(2) - F_1(1)}{1 - F_1(1)})(1 - \frac{F_2(3) - F_2(2)}{1 - F_2(2)})$

$= 1 - (1 - F_0(1))\frac{1 - F_1(2)}{1 - F_1(1)}\ \frac{1 - F_2(3)}{1 - F_2(2)}$

$\neq F_0(3)$ unless $F_1(1) = F_0(1)$, $F_2(2) = F_1(2)$ and $F_2(3) = F_0(3)$

Thus generalizing from these inequalities, neither of the synthesized *cdf*s G_C and G_A approximately equals the population *cdf* F_0 unless $F_i = F_0$ for all values of i which determine G_C and G_A. That is, neither function can be considered as an approximate replacement of a prevailing or current population lifetime *cdf*, F_0, unless the mortality patterns of previous years, specified by $F_1, F_2, F_3, \ldots, F_L$, have remained essentially unchanged for L years, the notional maximum human lifetime in the population under study.

It is usually assumed that, with the progression of time, there are general and widespread improvements in population living conditions such as public health and medical treatment, resulting in an overall shift to the right in the *cdf*s F_i. Alternatively, it could be anticipated that the expected value of the length of life derived from F_j is less than or equal to the expected value of the length of life derived from F_i, for j > i . (Remembering that from Section 1.5, the notation “j > i” means that calendar year j occurs before calendar year i.)

If this condition of “improvement” can be mathematically expressed as

$$\frac{F_i(x+y) - F_i(x)}{1 - F_i(x)} \le \frac{F_j(x+y) - F_j(x)}{1 - F_j(x)} \qquad \forall\ x, y \ge 0 \text{ and } j > i$$

it follows that

$$1 - \frac{F_i(x+y) - F_i(x)}{1 - F_i(x)} \ge 1 - \frac{F_j(x+y) - F_j(x)}{1 - F_j(x)} \qquad \forall\ x, y \ge 0 \text{ and } j > i$$

Thus for any value of i = 1,2,…L

$$1 - F_0(i)$$

$$= \left(1 - \frac{F_0(1) - F_0(0)}{1 - F_0(0)}\right)\left(1 - \frac{F_0(2) - F_0(1)}{1 - F_0(1)}\right)\ldots\left(1 - \frac{F_0(i) - F_0(i\text{-}1)}{1 - F_0(i\text{-}1)}\right) \qquad \text{Section 1.3.1}$$

$$\ge \left(1 - \frac{F_0(1) - F_0(0)}{1 - F_0(0)}\right)\left(1 - \frac{F_1(2) - F_1(1)}{1 - F_1(1)}\right)\ldots\left(1 - \frac{F_{i-1}(i) - F_{i-1}(i\text{-}1)}{1 - F_{i-1}(i\text{-}1)}\right) \qquad \text{from above}$$

$$\approx 1 - G_C(i).$$

The expected value of the length of life for the population of the calendar year with *cdf* F_0 is then

$$E[X \mid F_0]$$

$$= \int_0^L [1 - F_0(x)]\,dx \qquad \text{applying Section 1.3.3}$$

$$\approx \sum_{i=1}^{L} \tfrac{1}{2}[1 - F_0(i\text{-}1) + 1 - F_0(i)] \qquad \text{applying Section 1.4.1}$$

$$\geq \sum_{i=1}^{L} \tfrac{1}{2}[1 - G_C(i-1) + 1 - G_C(i)] \qquad \text{from above}$$

$$\approx E[X \mid G_C] \qquad \text{applying Section 1.4.1}$$

Thus, although G_C is only an approximation to F_0, and allowing for the further mathematical approximations in the above expressions, there is a tenuous argument that the expected value of the length of life derived from G_C is a lower bound for the expected value of the length of life of the currently prevailing population with *cdf* F_0. *ie* for the “true” average lifetime. A similar relationship can be established for the expectation calculated by using G_A but the details will not be given here.

An examination is made in Chapter 3 of the potential magnitude of the difference between $E[X \mid F_0]$ and $E[X \mid G_C]$ using South Australian data for illustration.

1.8 An alternative approach using the hazard estimator $u_k(k + j/2)$

Although not included in a conventional current life table analysis, two additional synthesized distribution functions defined by hazard functions are presented in this thesis for comparison of methodologies.

The quantities $u_0(\tfrac{1}{2})$, $u_1(1\tfrac{1}{2})$, $u_2(2\tfrac{1}{2}), \ldots, u_{100}(104\tfrac{1}{2})$ are used in the manner suggested by the approximation established in Section 1.3.2 to define a function H_C, analogous to G_C defined in Section 1.7, as

$$H_C(k+1) = 1 - \exp\left(- \sum_{i=0}^{k} u_i(i + \tfrac{1}{2}) \right), \text{ for } k=0,(1),104$$

Similarly, the quantities $u_0(\frac{1}{2})$, $u_1(1\frac{1}{2})$, $u_2(2\frac{1}{2})$, $u_3(3\frac{1}{2})$, $u_4(4\frac{1}{2})$, $u_5(7\frac{1}{2})$, $u_{10}(12\frac{1}{2})$,..., $u_{100}(102\frac{1}{2})$ are again used in the manner suggested by the approximation established in Section 1.3.2 to define a function H_A, analogous to G_A defined in Section 1.7, as

$H_A(k+1) = H_C(k+1)$, for k=0,(1),4

and

$$H_A(k+1) = 1 - \exp\left(-\sum_{i=0}^{4} u_i(i+\tfrac{1}{2}) - \sum_{i=5,(5)}^{k} 5u_i(i+\tfrac{5}{2})\right), \text{ for } k=5,(5),100$$

It is not productive to examine the mathematical relationship between $E[X \mid F_0]$ and $E[X \mid H_C]$ because H_C is defined in terms of mean value approximations to integrals, and not the integrals that specify F_0 itself.

1.9 Estimation of the lifetime distribution functions G_C, G_A, H_C and H_A

Each of these four functions, defined in Section 1.7 and Section 1.8, is based on the expected population size and the expected number of deaths for specified age groups. These expected values can be estimated from population vital statistics.

The expected population size for any age group is simply and directly estimated by the census population count taken in the calendar year C_0 for that age group.

The expected number of deaths for any age group is estimated, by convention, as the average of the observed number of deaths for that age group occurring in the year C_0, in the previous year C_1, and in the year following C_0, designated C_{+1} by extending the notation established in Section 1.5. This average value is used to reduce the effects of sampling variation, an issue that is discussed in detail in Section 1.11. The expected number of deaths occurring between the ages of birth and one year is used for illustration. With the notation of Section 1.5 and the result labelled OPTION 1(a), the expected number of deaths in this age range for the three consecutive calendar years are approximated by $N_1F_1(1)$, $N_0F_0(1)$ and $N_{+1}F_{+1}(1)$ respectively, and are equal assuming that $N_1 = N_0 = N_{+1}$ and $F_1 = F_0 = F_{+1}$. With this approximation, the average of the three observed values is an unbiased estimate of ${}_1D_0$, but with less variation than any of its components. Similar arguments apply to each age group.

The observed age-specific number of deaths and census population count are used in place of the expected values ${}_jD_k$ and ${}_jP_k$ in the expressions for ${}_jq_k$ and $u_k(k+j/2)$ given at the end of Section 1.5, and the resulting estimates are denoted by ${}_j\hat{q}_k$ and $\hat{u}(k+j/2)$ respectively.

The estimates ${}_j\hat{q}_k$ of ${}_jq_k$ are used in the expressions for G_C and G_A of Section 1.7 to produce estimated *cdf*s $\hat{G}_C(x)$, x=1,(1),104, and $\hat{G}_A(x)$, x=1,(1),5,(5),100. By definition, $\hat{G}_C(0)=\hat{G}_A(0)=0$, and for the purposes of this thesis, it is further defined that $\hat{G}_C(105)=\hat{G}_A(105)=1$. That is, the maximum attainable length of life (the quantity L of Section 1.3) is assumed to be 105 years. For notational convenience in other sections, this methodology is referred to as the q-method for complete and abridged life table analysis.

The estimates $\hat{u}(k+j/2)$ of $u_k(k+j/2)$ are used in the expressions for H_C and H_A of Section 1.8 to produce estimated distribution functions $\hat{H}_C(x)$, x=1,(1),104, and $\hat{H}_A(x)$, x=1,(1),5,(5),100. By definition, $\hat{H}_C(0)=\hat{H}_A(0)=0$, and for the purposes of this thesis, it is further defined that $\hat{H}_C(105)=\hat{H}_A(105)=1$. This methodology is referred to in other sections as the u-method for complete and abridged life table analysis.

A difficulty occurs in the estimation process when the census population count for any age group is zero, and hence the corresponding estimates of ${}_j\hat{q}_k$ and $\hat{u}(k+j/2)$ are undefined. These circumstances depend on the quality of the available data and the age structure of the population being studied, and most frequently occur for older age groups. The strategy used in this thesis to resolve this problem when it arises is discussed for $\hat{G}_C$, but it is also applied in principle to the other three estimated *cdf*s.

If ${}_1\hat{q}_k$ is undefined, but both ${}_1\hat{q}_{k-1}$ and ${}_1\hat{q}_{k+1}$ are defined, then the method of linear interpolation of Section 1.4.3 is used to impute a value for ${}_1\hat{q}_k$. Calculation of $\hat{G}_C(k+1)$ can then be made.

However, if ${}_1\hat{q}_k$ is undefined, and values of ${}_1\hat{q}_m$ are also undefined for all values of m greater than k, then imputed values for $\hat{G}_C(m+1)$, m= k,(1),100 are obtained by an extrapolation process.

As stated by Spiegelman [2] "a number of arbitrary methods have been used to supply mortality rates for this period of life. For practical purposes, any reasonable method is satisfactory, for the assumptions made will have only a small effect …" on, in particular, the expected value of length of life at birth. Elandt-Johnson [8] discusses a variety of methods that are used in these circumstances, detailing the mathematically sophisticated process of fitting a Gompertz distribution [20] to mortality data, as well as other methods that do not require the assumption of any specific analytic model. In Section 1.11 and Section 1.12, computationally intensive computer methods are presented and these are applied to an extensive range of data sets that are discussed in the sections of Chapter 2. Since the application is, by necessity, an automated process without the practicability of inspection and intervention for individual cases, a pragmatic and "fail safe" method of extrapolation is needed and has been adopted for this thesis. The average of the last three valid estimates, ${}_1\hat{q}_{k-3}$, ${}_1\hat{q}_{k-2}$ and ${}_1\hat{q}_{k-1}$, is combined with an assumed value of ${}_1q_{105} = 1$ to linearly interpolate a value for ${}_1\hat{q}_k$ by applying the method of Section 1.4.3. Calculation of $\hat{G}_C(k+1)$ can then be made.

The implications of this strategy are considered within the context of specific applications to data described in the sections of Chapter 2, acknowledging the warning given by Elandt-Johnson that " the uncertainties inherent in extrapolation should always be born in mind".

1.10 Estimation of current expected life

An approximation to the expected value of the length of life at birth for the distribution defined by G_C is determined by specializing the general results from Section 1.3.3 with L=105. Thus

$E[X \mid G_C]$

$$= \int_0^{105} [1 - G_C(x)]\, dx$$

$$\approx \int_0^{105} [1 - \hat{G}_C(x)]\, dx$$

$$= E[X \mid \hat{G}_C]$$

$$= \sum_{i=1}^{105} \int_{i-1}^{i} [1 - \hat{G}_C(x)]\, dx$$

$$\approx \tfrac{1}{2} \sum_{i=1}^{105} [2 - \hat{G}_C(i\text{-}1) - \hat{G}_C(i)] \qquad \text{applying Section 1.4.1}$$

Likewise, $E[X \mid H_C] \approx E[X \mid \hat{H}_C] \approx \tfrac{1}{2} \sum_{i=1}^{105} [2 - \hat{H}_C(i\text{-}1) - \hat{H}_C(i)]$

Approximations to $E[X \mid G_A]$ and $E[X \mid H_A]$ are established in a similar way, but with appropriate modifications to reflect the two different step sizes of one and five years used to determine the function values. Hence

$$E[X \mid G_A] \approx E[X \mid \hat{G}_A] \approx \tfrac{1}{2} \sum_{i=1}^{5} [2 - \hat{G}_A(i\text{-}1) - \hat{G}_A(i)] + \tfrac{5}{2} \sum_{i=10,(5)}^{105} [2 - \hat{G}_A(i\text{-}5) - \hat{G}_A(i)]$$

and

$$E[X \mid H_A] \approx E[X \mid \hat{H}_A] \approx \tfrac{1}{2} \sum_{i=1}^{5} [2 - \hat{H}_A(i\text{-}1) - \hat{H}_A(i)] + \tfrac{5}{2} \sum_{i=10,(5)}^{105} [2 - \hat{H}_A(i\text{-}5) - \hat{H}_A(i)]$$

These estimated values are collectively referred to as estimates of current expected life. A specific numerical estimate is often qualified by the type of current life table on which it is based, and by the method of calculation. (*i.e.* complete or abridged life table analysis; q-method or u-method)

1.11 Estimation of the variance of the estimate of current expected life

The quantification of the random error in the estimated expectations presented in Section 1.10 has received little attention in the literature of current life table methodology. An actuarial perspective is offered by Benjamin & Haycocks [10]. While certainly acknowledging the sampling error of derived rates and the possibility of some role for statistical confidence intervals, they present the actuarial viewpoint by stating that these "could not be of any practical value to actuarial calculations which require the application at any age of a specific rate of mortality and not a range of rates". For actuarial purposes, cautious and careful graduation, or smoothing of the data, is the generally preferred option to produce current life tables with regularity of change in mortality rates with age. From a statistical perspective, this attitude masks the variability of the data on which the current life table is based, and consequently gives no indication of the statistical accuracy of any derived summary statistic, in particular the expected value of the length of life at birth.

It might be considered that estimation error is negligible because of the large sample sizes on which current life tables are often based. However this is not always the case. For example, current life tables may be constructed for a demographic assessment of regional areas with small populations, and using the estimated expected values of length of life at birth as summary statistics. Any conclusions concerning regions would then depend on the variability of the estimates, which could not necessarily be discounted by virtue of sample size. Selvin [11] presents this view succinctly.

A leading proponent for the quantification of statistical error in the many varieties of estimates derived from current life table analysis is Chin Long Chiang. He describes the stochastic nature of the life table in a series of scientific publications in the 1960's [21], [22] [23] and in a classic textbook in 1984 [14]. In particular, he considers the estimate of the expected value of length of life at birth as a function of the random variables ${}_1\hat{q}_k$, each of which is assumed to be binomially distributed and hence with a known variance [24]. To this function, he applies a standard statistical procedure for a general function of random variables; the details of which can be found, for example, in Kendall [25]. This procedure is also known as the method of statistical differentials or the delta method [26], and caution is always recommended in any particular application of this method. Only the first-order linear terms of a Taylor series approximation to the pertinent function are used, and this level of approximation may not be sufficiently accurate in any given situation. Chiang's methodology, occasionally acknowledged and referenced in the literature, has not been routinely adopted for published official current life tables. Chiang's method is implemented in the IMSL library of FORTRAN functions and subroutines [27] (and unfortunately the routine is defective), but it is not available in any other widely distributed statistical software package. The arithmetical calculation is not particularly onerous or difficult, being a recursive summation of various basic current life table quantities, and Chiang provides FORTRAN coding for the calculations in his book. It may be that Chiang conveys the impression that statistical variation can be discounted through his choice of the particular numerical examples that are used for the illustration of his technique. For example, he gives the expected value of length of life at birth for white males in the United States in 1955 as 67.3 years with an estimated standard error of .0181 years, or approximately one week. This small error reflects the large total size of his study population of approximately 73 million and is not necessarily typical of other current life table applications.

In this thesis, the modern statistical procedure known as bootstrapping is used to quantify the statistical variation in the estimates of the expected value of length of life at birth that are presented in Chapter 2. Efron formalised the bootstrap principle in 1979 [28] and the method is clearly and extensively discussed, and the statistical properties justified, in the textbook by Efron & Tibshirani [29]. In the introduction to Chapter 2 of this book the authors say: "The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates. The basic idea behind the bootstrap is very simple, and goes back at least two centuries". In essence, a large number of samples of values are (pseudo-)randomly produced using a computer program and a generating process based on an observed sample of data. The quantity of interest is calculated from each generated (pseudo-)random sample, and the variation between these (pseudo-)random estimates is used to infer an error estimate for the particular estimate calculated from the observed sample of data.

The implementation of the bootstrap methodology is described in detail for the estimated expectation $E[X \mid \hat{G}_C]$. Since the application similarly applies to the expectations based on the other three estimated *cdf*s $\hat{G}_A$, $\hat{H}_C$ and $\hat{H}_A$, it will not be explicitly described for these functions. Preliminary to this description, however, some necessary definitions and results are established in Section 1.11.1 and Section 1.11.2.

1.11.1 The Triangular Distribution: T(a,m,b)

A random variable X with a general triangular distribution, defined on the range [a,b] with a unique mode at m, $a \le m \le b$, has a *pdf* t(x) given by

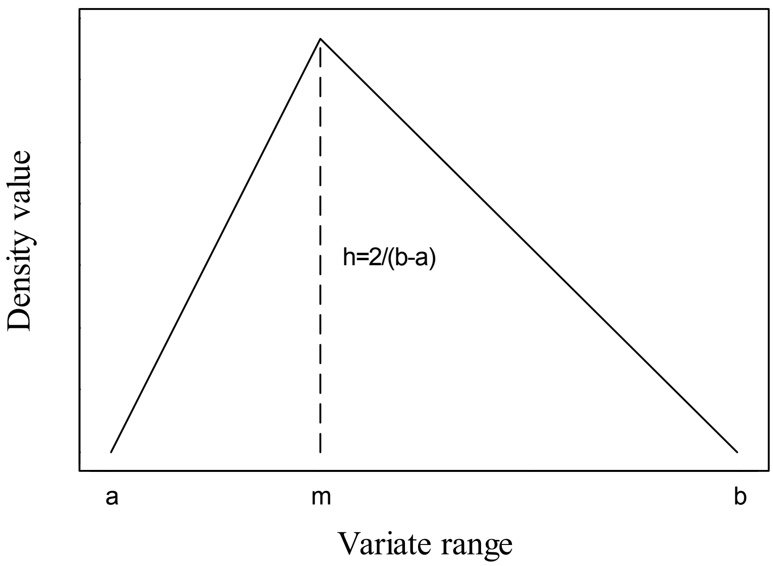
t(x) = h (x-a) / (m-a) = 2(x-a) / [(b-a)(m-a)] for $a \le x \le m$

and

t(x) = h (b-x) / (b-m) = 2(b-x) / [(b-a)(b-m)] for $m \le x \le b$

The notation **T**(a,m,b) used in this thesis to denote the triangular distribution and the density is illustrated in Figure 1.11.1.

Figure 1.11.1: Triangular distribution T(a,m,b)



The *cdf* of X, T(x), is given by

$T(x) = h\,(x\text{-}a)^2 / 2(m\text{-}a) = (x\text{-}a)^2 / [(b\text{-}a)(m\text{-}a)]$ for $a \le x \le m$

and

$T(x) = 1 - h\,(b\text{-}x)^2 / 2(b\text{-}m) = 1 - (b\text{-}x)^2 / [(b\text{-}a)(b\text{-}m)]$ for $m \le x \le b$

The expected value of X is $E[X] = \frac{1}{3}[\,m + (a+b)]$; and E[X] = m for the special case of the symmetric triangular distribution, in which $m = \frac{1}{2}(a+b)$.

Further reference to the triangular distribution can be found in standard texts: *see* Johnson [20] or Kotz [26].

A (pseudo-)random value v can be generated from **T**(a,m,b) by a simple and direct method using the inverse function of the *cdf* T(x); *see* Kennedy & Gentle [30], for example, for a detailed discussion of this general procedure. In particular, with T(m) = (m-a)/(b-a),

(a) A (pseudo-)random value u is generated from the Uniform distribution [20] on the range [0,1] by using an algorithm that has statistically validated properties of "randomness". The algorithm of Wichmann & Hill [31] [32] is used in this thesis.

(b) If $u \le T(m)$, then $v = a + \sqrt{u(m\text{-}a)(b\text{-}a)}$.

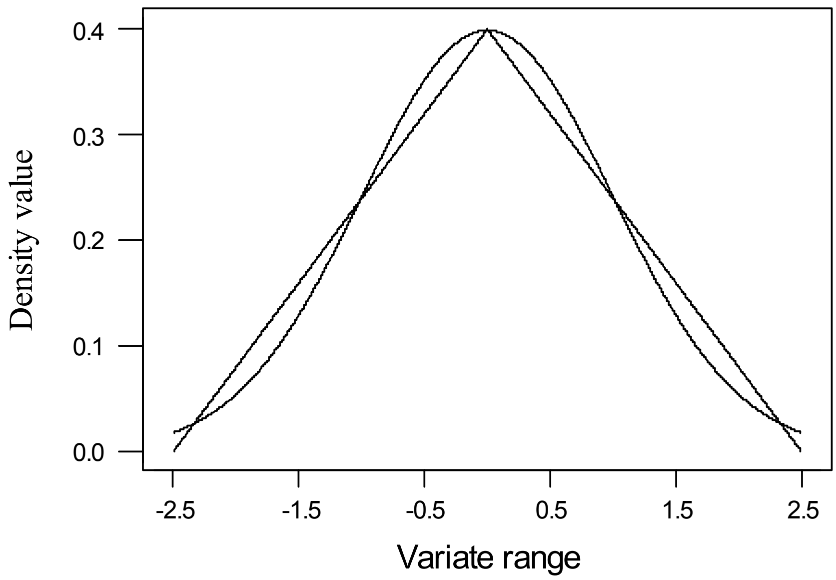
(c) If $u > T(m)$, then $v = b - \sqrt{(1\text{-}u)(b\text{-}m)(b\text{-}a)}$.

For a given range [a,b], there is a family of triangular distributions defined by all modal values $m = \{\, m : a \le m \le b \}$. The two distributions with the extreme modal values of this family, **T**(a,a,b) and **T**(a,b,b), have expected values of $\frac{1}{3}(2a+b)$ and $\frac{1}{3}(2b+a)$ respectively. For any value c within this range, $\frac{1}{3}(2a+b) \le c \le \frac{1}{3}(2b+a)$, there is a triangular distribution **T**(a, 3c-(a+b), b) with expected value E[X | **T**(a, 3c-(a+b), b)] = c.

1.11.2 Approximation of the Standard Normal distribution by the Triangular

The *pdf* of the Standard Normal distribution, with mean zero and variance one, N(0,1), is shown with the *pdf* of the triangular distribution **T**(-2.5,0,2.5) in Figure 1.11.2. Over this range [-2.5, 2.5], the symmetric triangular distribution encompasses approximately 98% of the probability mass of N(0,1), and implies that **T**(-2.5,0,2.5) is a useful working approximation to N(0,1). The triangular distribution is one of five distributions considered by Chew [33] as substitution distributions for the Standard Normal distribution for the purpose of "mathematical convenience".

Figure 1.11.2: Comparison of *pdf*s N(0,1) & T(-2.5,0,2.5)



1.11.3 Bootstrap estimation of the variance of the estimate of current expected life

The estimate $E[X \mid \hat{G}_C]$ is a function of the random variables ${}_1\hat{q}_k$, k=0,(1),104. Suppressing the subscript notation introduced in Section 1.5 for convenience and clarity in this section, the basic steps of the bootstrap algorithm implemented in this thesis are:

1. With D being the observed number of deaths and P being the observed population size in an age group, then D is assumed to be binomially distributed [34]. This assumption is also the basis of the method proposed by Chiang. Using the normal approximation to the binomial distribution, two values of $q = D / (P + \frac{1}{2} D)$ are found corresponding to the 1% and 99% points of the distribution (constrained in application to be at least zero and at most one).
2. These two percentile values are expressed as the equivalent lower and upper limits for the distribution of the number of deaths, D_1 and D_2 respectively. The nominal normality implied as the distribution for the number of deaths as a consequence of the assumptions in step 1 is approximated by $\mathbf{T}(D_1, D, D_2)$, following from Section 1.11.2.
3. A (pseudo-) random number of deaths d is generated from $\mathbf{T}(D_1, D, D_2)$ and a (pseudo-)random value of $q = d / (P + \frac{1}{2} d)$ is calculated.
4. Steps 1 - 3 are repeated for each implied value of k, and a (pseudo-)random value of $E[X \mid G_C\{\text{bootstrap sample}\}]$ is calculated using the set of 105 (pseudo-)random values of q.
5. Steps 1 - 4 are repeated B times, called bootstrap replications, and the arithmetic mean and standard deviation are calculated from the B (pseudo-) random values of $E[X \mid G_C\{\text{bootstrap sample}\}]$. These B values are ordered, and the median value, and the lower and upper bounds of a central 90% confidence interval, are calculated.

The standard deviation obtained at step 5 is the bootstrap error of the estimate $E[X \mid \hat{G}_C]$ calculated from the observed sample. The variation in the sample estimate $E[X \mid \hat{G}_C]$ is alternatively expressed through the bootstrap 90% central confidence interval. The arithmetic mean obtained at step 5 is an unbiased estimate of the sample estimate $E[X \mid \hat{G}_C]$, and these two values are essentially numerically equal in a correctly implemented algorithm using a suitably large value for B. Since distributions are assumed and parameters estimated at steps 1 and 2, the procedure is termed "parametric bootstrap" by Effron [29].

Effron also states that from his experience "Very seldom are more than B = 200 replications needed for estimating a standard error. (Much bigger values of B are required for bootstrap confidence intervals)." Naturally computation time increases with B, but since computing time was not an issue for this thesis, B=5001 was chosen. The generation of non-integer values for the number of deaths at step 3 by using $\mathbf{T}(D_1, D, D_2)$ is, at most, a concept problem and can be viewed as a means-to-an-end to produce values of q that have appropriately induced variability. It is the property of "mathematical convenience" discussed by Chew [33]. Moreover, the use of $\mathbf{T}(D_1, D, D_2)$ automatically and simply restricts (pseudo-)randomly generated values of the number of deaths to an appropriately constrained range for computational convenience.

1.11.4 An additional source of variation in the estimate of current expected life

In the bootstrap algorithm described in Section 1.11.3, it is assumed that the population sizes for each of the age groups, generically denoted P, are fixed values and are not subject to sampling variation. However, since in practice they are obtained as census counts, they too can also be considered as potentially subject to some sort of error process. While an argument is made in Section 1.11.3 step 2 to justify the distribution of the number of deaths, a distribution for the population size has to be imposed by external considerations.

The algorithm described above is expanded to include at step 3

…and a (pseudo-) random population size p is generated from $\mathbf{T}(P_1, P, P_2)$…

with p subsequently used in place of P to calculate q.

The choice of values for P_1 and P_2 is made in the context of the circumstances of the census producing the age-specific population counts. (*see* Section 2.5) Clearly, if $P_1 = P = P_2$, every value from the degenerate distribution $\mathbf{T}(P,P,P)$ is equal to P and the situation becomes that described in Section 1.11.3. The other steps in the algorithm are unchanged.

1.11.5 A robustness determination of the estimate of current expected life

In Section 1.11.3 and Section 1.11.4, variation in $E[X \mid \hat{G}_C]$ is assumed to occur through variation in the age-specific number of deaths, and, additionally, through variation in the age-specific census population counts. Both these sources of variation are considered random and unbiased. However the literature of population studies often casts doubts on these assumptions and for a variety of reasons suggests that reported figures are undercounts of the true situation. The bootstrap algorithm of Section 1.11.3 can be expanded to examine the potential consequences of under-reporting of either the number of deaths, or the population size, or both, on the magnitude and variation of $E[X \mid \hat{G}_C]$.

Again using the generic notation established in Section 1.11.3, the distribution of the number of deaths can be generalized as $\mathbf{T}(D_1, \delta_1 D, D_2)$ where $\delta_1 \geq 1$. With $\delta_1 = 1$, the distribution becomes that described in Section 1.11.3 step 3, and corresponds to the situation of random variation around the observed number of deaths D. However, suppose it is considered that, because of under-reporting, the true number of deaths is larger than the observed number and can be reasonably represented by some fractional multiple of D, $\delta_0 D$ where $\delta_0 > 1$. Using the relationship between the mode and the expected value of a triangular distribution established in Section 1.11.1, a value of $\delta_1 = [3\,\delta_0 D - (D_1 + D_2)] / D$ defines an asymmetric triangular distribution $\mathbf{T}(D_1, \delta_1 D, D_2)$ that has an expected value $\delta_0 D$. Thus, through a specification of the parameter δ_0 and the associated derived value for δ_1, a distribution is produced that has an expected value larger by a nominated factor than the observed number of deaths. *i.e.* an "upward" shift of the distribution to reflect the concept of under-reporting. The choice of the size of the multiple δ_0 of D is dependent on the particular characteristics of the population under study, and reflects the degree of under-reporting of deaths that is thought to have occurred, for whatever reasons.

Similarly, the distribution of the size of the population can be generalized and reformulated as $\mathbf{T}((1-\delta_2)\delta_3 P,\ \delta_3 P,\ (1+\delta_2)\delta_3 P)$ where $0 \leq \delta_2 \leq 1$ and $\delta_3 \geq 1$. Suitable choice of values for δ_2 and δ_3 produce distributions for the size of the population corresponding to a variety of circumstances: an invariant value for P requires $\delta_2 = 0$ and $\delta_3 = 1$; a distribution for P subject only to random variation requires $\delta_2 > 0$ and $\delta_3 = 1$; while a distribution for P reflecting both under-reporting of population size and with random variation requires $\delta_2 > 0$ and $\delta_3 > 1$.

Steps 2 and 3 of the bootstrap algorithm of Section 1.11.3 are expanded to include generalizations for the two distributions:

the number of deaths d as $\mathbf{T}(D_1, 3\,\delta_0 D - (D_1 + D_2), D_2)$

the size of the population p as $\mathbf{T}(\,(1-\delta_2)\delta_3 P,\ \delta_3 P,\ (1+\delta_2)\delta_3 P\,)$

With this extended formulation, the objective of the values assigned to the row vector $\boldsymbol{\delta} = (\delta_0, \delta_2, \delta_3)$ is to reproduce the assumed distributional properties of the data from which the sample estimate $E[X \mid \hat{G}_C]$ is derived. If either $\delta_0 > 1$ or $\delta_3 > 1$, the algorithm might more properly be described as a Monte Carlo simulation [30]. For choices of parameter values in this range, the bootstrap sample mean calculated at step 5 will no longer be necessarily equal to the sample estimate $E[X \mid \hat{G}_C]$, and the magnitude of the difference in these two values will indicate the extent of the bias in the sample estimate that can be produced by the assumed type of under-reporting.

The value of $\boldsymbol{\delta}$ that is appropriate to any particular application is most probably not constant across the age groups used to calculate $E[X \mid \hat{G}_C]$, and perhaps should be indexed to indicate this fact. However, for practical reasons that are discussed in detail in Chapter 2, a common value of $\boldsymbol{\delta}$ across age groups is assumed for bootstrap estimation in this thesis.

1.12 A comparison of methodologies

The estimation of the distribution functions described in Section 1.9, the calculation of the various estimates of expected value of length of life at birth described in Section 1.10, and the bootstrap procedure to examine the variability and robustness of these estimates described in Section 1.11; have been implemented in a FORTRAN computer program for use in this thesis. The various operational requirements, features and output of this program are described in Section 2.4 of Chapter 2.

The 1955 data for white males in the United States, analysed by Chiang [22], is used for comparison purposes and to some extent as a validation of the coding of the thesis computer program. The results published by Chiang are shown in Table 1.12, along with values calculated for this thesis using his data. The two expectations shown in row 1 are equal, after rounding to the first decimal place specified by Chiang. This agreement is anticipated as each expectation is calculated the by independent application of standard current life table methodology, although Chiang does use the adjustment to the ${}_jq_k$ -values discussed in Section 1.6 and each method uses a slightly different method of numerical integration.

Table 1.12: Comparison of Chiang and Thesis results for US White Males 1955

Estimate		Method	
		Chiang	Thesis
1	Current expected life	67.3	67.2622
2	Standard error	.0181	
	Bootstrap		
3	Arithmetic mean		67.2624
4	Standard deviation		.0187
5	5th Percentile		67.2313
6	Median		67.2626
7	95th Percentile		67.2935
8	Range error =(95th -5th)/3.29		.0189

The results from the thesis method shown in row 1 and row 3 are almost identical to four decimal places, the specified accuracy of Chiang. This agreement strongly indicates that the bootstrap algorithm has been implemented correctly into the thesis computer program. The error estimates from the two methods, shown in row 2 and row 4, are again almost identical. This level of agreement, while not necessarily surprising, is reassuring in view of the comments made in Section 1.11 about the delta method that forms the basis for Chiang's technique. Another estimate of the error of the estimate of the expected value of length of life at birth, derived from percentile values [24], is shown in row 8. It also agrees very closely with the other two estimates of error.

The overall agreement in these results, both between and within methods, is supportive that the thesis computer program functions correctly. Other application to data presented in Chapter 2 also support this conclusion, at least with respect to the numerical calculation of estimates of the expected value of length of life at birth. The validity of the process of error estimation relies on the comparisons presented in Table 1.12, exhaustive checking of the FORTRAN code, and, ultimately, the theory of the bootstrap.

CHAPTER 2: APPLICATION TO SOUTH AUSTRALIAN DATA

2.1 The sources of data for the estimation of current expected life for South Australia

Extensive population and mortality information has been recorded in South Australia since it was first settled as a British colony in 1836. These data essentially refer to people of British and Western European ancestry. Indigenous Aborigines were excluded from official records until the middle of the 20th century, and there has been minimal immigration from non-European countries until the latter part of the 20th century.

The collection of population census data was initially the responsibility of the Governor of the colony and of the Colonial Secretary, and became that of the Chief Secretary's office with the evolution of responsible government and the establishment of a parliamentary Legislative Council. Eleven recognised colonial censuses were taken in the years 1844, 1846, 1851, 1855, 1861, 1866, 1871, 1876, 1881, 1891 and 1901. An enumeration or "mustering" of the population was made in 1841 but this is not recognised as an official census. The results of a census held on April 1st 1860 (April Fools Day) are usually disregarded because of, as reported in the Parliamentary Papers of 1861, the generally considered "absence of a portion of the population and other disturbing influences". An official review and discussion of all colonial era censuses, for South Australia and the other Australian colonies, can be found in the *Statistician's Report on the Census of the Commonwealth of Australia 1911*. Commentary on colonial censuses of Australia can be found in Camm [35], and South Australian colonial censuses are discussed in detail in Stevenson [36].

The results of the colonial censuses are available in a variety of official publications, not necessarily mutually exclusive. Sources used to obtain the gender and age-specific population data required for this thesis are:

- British House of Commons Parliamentary Paper, 1843, V32, No 505 (BHCPP). This is an early report on the progress of the colonisation of South Australia and includes the results of the "mustering" of 1841.
- Statistics of South Australia 1845-1846 (SSS) and Statistical Returns of South Australia 1841-1858 (SRSA). These are the earliest colonial government publications of population and other statistical information. Some original manuscripts are held in

the Mortlock Collection, State Library of South Australia, with photocopies available for general use.

- Statistical Register of South Australia 1859-1975 (SRSA). This series continued and expanded the range of statistical information contained in Statistical Returns. The series was initially published by the Chief Secretary's office until the Federation of Australia in 1901, and thereafter by the Australian Bureau of Statistics (ABS ; called the Commonwealth Bureau of Census and Statistics from 1905-1974). These annual reports are available in the reference section of the Barr-Smith Library of the University of Adelaide, in the State Library of South Australia, and elsewhere.
- Census reports included as appendices to Proceedings of the Parliament of South Australia 1855-1901 (SAPP).

The data relating to any given census is often, but not necessarily, presented in two or more of the above sources. Detailed indexes to the sources of a variety of Australian colonial statistics are contained in Miller [37], and, in particular for South Australia, in Peake [38] and in occasional papers prepared for the ABS by Pitt [39] and Pennock [40]. Much of the material referred to in these sources has been incorporated into an extensive set of micro-fiche produced by the ABS [41]. Unfortunately the readability of the locally available copy held by the Barr-Smith Library of the University of Adelaide has deteriorated in clarity and thus was of limited practical use for this thesis.

After the Federation of the Australian colonies in 1901, the ABS became responsible for conducting national censuses. Thirteen censuses have been taken by the ABS for the years 1911, 1921, 1933, 1947, 1954, 1961, 1966, 1971, 1976, 1981, 1986, 1991 and 1996. The results of these censuses are well documented and readily available in publications of the ABS, initially presented as results obtained for a specific census but often repeated in other reports requiring gender and age-specific population data. Additional detail for higher ages was obtained from the ABS by privately commissioned single-year age tabulations for the censuses of 1986, 1991 and 1996.

An Act of Parliament (*5 Victoria*, No 13, 1842) established the Registry of Births, Deaths & Marriages (RBDM) in South Australia. From that time, it became a legal requirement that each birth be notified in writing to the local Deputy Registrar within forty-two days after birth, and each death within ten days of death. The responsibility for registration of deaths continued as a State function after Federation. Sources used to obtain the gender and age-specific mortality data required for this thesis are:

- "Index to deaths prior to compulsory registration 1802-1842". This compilation, produced by the RBDM in 1959, summarises
  (a) the burial records of Holy Trinity Church from February 9th 1837 until March 30th 1842. Holy Trinity was the pioneer church of the Church of England in South Australia.
  (b) the burial records of West Terrace Cemetery from July 6th 1840 until March 30th 1842. This was the first public cemetery in South Australia.
- Archive folio manuscripts of the RBDM held by South Australian State Records. These volumes, beginning with cursive handwritten entries in April 1842 and continuously accumulated since that time, are the official death records for the Colony, Province, and State of South Australia. Permission to access this material was given by the Registrar of the RBDM in 1998.
- The South Australian Government Gazette, 1839-continuing (SAGG). Issued weekly, this series contains the irregularly published annual reports of the RBDM for colonial South Australia.
- Statistical Register of South Australia, 1859-1975.
- Privately commissioned single-year age tabulations for each gender produced by the ABS from archived data, for the years 1976-82, 1985-87, 1990-92 and 1995-97.
- Australian Demography Bulletin, 1906-continuing. (Initially known as the Australian Population and Vital Statistics Bulletin) Published periodically by the ABS.
- Deaths, South Australia, 1969-1989. Published annually by the ABS.
- Demography, South Australia, 1990-continuing. Published annually by the ABS.

The sources of the specific data analysed in Section 2.6 to Section 2.15 is often indicated by using the alternate parenthesised form of the above titles and the year of publication.

2.2 The computing environment of the computer programs used for this thesis

The implementation of the methods described in Chapter 1 to South Australian data has been through the usage of computer files and programs appropriate to the Microsoft Windows ™ operating system. The notation "<< x.xx >>" is used throughout this thesis to indicate a computer file named " x.xx " compatible with this environment, and with the standard suffix convention of " .xx " denoting a file suitable for specific types of application. The large number of population data files, original computer programs and resultant output files produced by these programs have been copied onto a CD-rom that is included as an appendix to this thesis. Most files are of the type " .txt " and can be viewed using basic Microsoft products, *e.g.* the straightforward text editor Notepad.

2.3 The naming and structure of data files

The gender and age-specific population and mortality data compiled for South Australia from the sources described in Section 2.1 have been entered into one of fifty text files (*included* CD-rom). These data files are generically named << *gyyyy*.txt >>, with the naming connotation of

g indicating gender, and is either M (for male data) or F (for female data)

yyyy indicating the calendar census year of the population count, and the central calendar year for the mortality data.

The content of the data files is specifically arranged for direct input into the thesis computer program described in Section 2.4. For each data file, the nine columns of values are:

column 1 : an alpha-numeric character variable, reiterating the filename
column 2 : a lower single-year age-group boundary value
column 3 : an upper single-year age-group boundary value
column 4 : the mid-point of the single-year age-group (for a complete current life table)
column 5 : the mid-point of a five-year age-group (for an abridged current life table)
column 6 : the single-year age-group population count for the census year *yyyy*
column 7 : the single-year age-group number of deaths for the year "*yyyy*-1"
column 8 : the single-year age-group number of deaths for the year *yyyy*
column 9 : the single-year age-group number of deaths for the year "*yyyy*+1"

An extract from << M1996.txt >>, listing the male population count from the 1996 census, and the male mortality data for the years 1995, 1996 and 1997, is shown in Table 2.3.1.

Table 2.3.1: Extract from the data file << M1996.txt >>

'M1996'	0	1	0.5	0.5	9387.0	73.0	57.0	53.0
'M1996'	1	2	1.5	1.5	9552.0	2.0	3.0	4.0
'M1996'	2	3	2.5	2.5	9943.0	1.0	1.0	7.0
'M1996'	3	4	3.5	3.5	9926.0	4.0	4.0	5.0
'M1996'	4	5	4.5	4.5	10212.0	4.0	4.0	1.0
'M1996'	5	6	5.5	7.5	10086.0	3.0	2.0	3.0
'M1996'	6	7	6.5	7.5	10151.0	0.0	1.0	1.0
'M1996'	7	8	7.5	7.5	10164.0	2.0	1.0	2.0
'M1996'	8	9	8.5	7.5	10014.0	1.0	2.0	3.0
'M1996'	9	10	9.5	7.5	9907.0	1.0	0.0	0.0
'M1996'	10	11	10.5	12.5	10276.0	1.0	0.0	0.0
'M1996'	11	12	11.5	12.5	10293.0	0.0	2.0	1.0
'M1996'	12	13	12.5	12.5	10530.0	0.0	1.0	1.0
'M1996'	13	14	13.5	12.5	10347.0	2.0	5.0	1.0
'M1996'	14	15	14.5	12.5	10088.0	2.0	1.0	3.0

The features and quality of the population and mortality data recorded in each individual data file are discussed in conjunction with the results of the current life table analysis of that data file; *see* Section 2.6 to Section 2.15.

2.4 The thesis FORTRAN computer program: << current.f >>

The source code of the thesis FORTRAN computer program for current life table analysis is contained in the file << current.f >> (*included* CD-rom). The executable form of this program, << current.exe >>, (*included* CD-rom) was created by the f77 compiler provided by Absoft Pro Fortran for Windows [42]. The program is executed in an MS-DOS window by the command line statement " current < currentin.txt ".

The standard text file << currentin.txt >> is a listing of specific data file names, using the file naming convention established in Section 2.3, for which a current life table analysis is required.

The standard text file << bootstrap.txt >> is an additional file that is referenced when << current.exe >> is executed. Values are assigned to the parameters B and $\boldsymbol{\delta}=(\delta_0, \delta_2, \delta_3)$, defined in Section 1.11.3 and Section 1.11.5, through the sets of four numbers included in this file. Each line of << bootstrap.txt >> produces a corresponding bootstrap analysis. The choice of values for the specific version of << bootstrap.txt >> (*included* CD-rom) used to calculate results reported in this thesis is discussed in detail in Section 2.5.

The execution of << current.exe >> creates two text files as output for each data file << *gyyyy*.txt >> listed in << currentin.txt >>. These output files have program-generated names of the form << CELQ*gyyyy*.txt >> and << CELU*gyyyy*.txt >> (*included* CD-rom), with the naming connotation that

CEL — is an abbreviation indicating estimates of C(urrent) E(xpected) L(ife) resulting from current life table and bootstrap analyses.

Q and U — is an abbreviation indicating a current life table analysis using either the q-method or the u-method (*see* Section 1.9).

Each output file has two sections that display the results derived from a complete current life table analysis and an abridged current life table analysis (*see* Section 1.7). Both sections contain all bootstrap analyses specified by << bootstrap.txt >>. A typical example of an output file, << CELQM1996.txt >>, is displayed and described in detail in Section 2.6.

The computational time required to analyse any given data file is a function of the hardware of the computer used for the analysis, and the number and size of the bootstrap analyses specified by <<bootstrap.txt>>. The computing arrangement described here allows for a sequence of analyses to be progressively executed in a dedicated window application.

2.5 The specification of the bootstrap analyses: << bootstrap.txt >>

It is generally recognized that there are three types of error in population and mortality counts, arising from

- individuals not included in the count at all (undercount)
- individuals being counted, but not stating an age (unstated)
- individuals being counted, but stating an incorrect age (misstated)

The first Australian Statistician, G. H. Knibbs, discussed these features in the *Statistician's Report on the Census of 1911*, and this discussion was continued by his successor, C. H. Wickens, in the *Statistician's Report on the Census of 1921*. The estimated percentage error in the national population counts, combined over gender and over all ages, is shown in Table 2.5.1.

Table 2.5.1: Estimates of error rates in the Censuses of 1911 and 1921

Type of Error	ABS Census 1911	ABS Census 1921
Uncounted	-	-
Unstated	.53%	.26%
Misstated	1.18%	1.15%

In a subsequent paper on Australian mortality in 1930, Wickens [43] expresses the opinion that “in the census enumeration itself complete accuracy cannot be guaranteed, but…errors of enumeration are comparatively small”. He also states that “ Registrations of births and deaths are believed to attain a high degree of completeness in Australia, and where tests have been possible, confirmation of this belief has been obtained, but even here there are almost certainly some omissions” and that “these [death] figures involved similar omissions and misstatements to those in evidence in the census results, but there were relatively fewer omissions and at many ages relatively more misstatements”.

In the *Year Book of the Commonwealth of Australia* 1936, the Australian Statistician R. Wilson states “…the officials responsible for the census of 1933 feel hopeful that in the population return of that census, and the related death returns of 1932-34, a relatively high degree of reliability has now been reached”. As had previous Australian Statisticians, Wilson also commented on the “psychological peculiarity” present in the age distributions, with excess counts for ages ending in 0 and 5, amongst others. Of particular relevance to current life table analyses, he further states that “The fact that this peculiarity appears in both numerator [*i.e.* deaths also] and denominator of the fraction from which q_x is obtained assists in reducing the disturbing effect, and it is still further diminished by the [age] grouping..”.

The *Statistician’s Report on the Census of 1961* provides figures that show an error rate for unstated age of approximately .4%, averaged over both genders and all ages, and the six national censuses taken by the ABS up to and including 1961.

Since 1976, the ABS has monitored the accuracy of census returns using post-enumeration surveys of increasing methodological complexity that have provided estimates of the size of the undercount of the census. The results of these studies have been presented in ABS *Census 86: Data Quality Undercount* and ABS *Census 91: Data Quality Undercount*.

The undercount rate for South Australia, averaged over both genders and all ages, was estimated as 1.6% for each of the censuses taken in 1981, 1986 and 1991. These rates were amongst the lowest nationally.

A detailed examination of national undercount rates for the 1996 census was published in 1999 in the ABS Demography working paper No 99/4, which indicates overall undercount rates of approximately 2.0% for males and 1.2% for females. A distribution of undercount rates using five-year age groupings is also provided for each gender. Generally, males have a higher undercount rate than females, with maximums of approximately 4% and 3% occurring between the ages of 20 and 30 years, respectively.

There is little direct evidence concerning the accuracy of the colonial censuses taken in South Australia prior to Federation. Stevenson [36] has produced a resource paper describing the history and development of the census process in Australasia with special reference to colonial South Australia. He reviews the detailed organization that was used to minimize undercounting and misstatement through “collectors…chosen for the local knowledge and reliability…and check through the completed forms for omissions or obvious errors”. He states “it was believed that the organization of the distribution and collection of householder schedules was so tight and thorough, as to preclude any ordinary grounds for error”. There is no quantification of the level of this belief.

Unsubstantiated estimates of undercount were sometimes provided for South Australian census results. For example, for the census of 1861 reported in SAPP 1862, a total population count of 126,830 persons is given “To which may be added for migratory and unenumerated persons, 1,170 - a low count considering…”. These figures produce an undercount rate of approximately 1%, which is a better result than that achieved by the ABS in South Australia in the 1980’s. A comparison with the results of the disregarded census of 1860 can perhaps be used to infer an upper bound for census undercount at that time. A total population count of 117,967 persons was reported in the 1860 census, which gives an approximate ratio of 1.08 of the “true” 1861 result to the “false” 1860 result. It would be perhaps anticipated that an undercount error rate would be less for a diligent and properly conducted census than was reportedly the case in 1860.

Inferences may be made to South Australian censuses from the more extensive commentaries available for other Australian colonies. Camm [35] reviews many aspects of colonial census taking, and reports that “the five to six percent added to the first census of

New South Wales was in line with the inaccuracy which is thought to have been present in the first English census of 1801". Camm also reports studies undertaken in Queensland which concluded that for 1869 "…two and a half per cent may be fairly added…", reduced to 1.5% in 1886 "in the light of improved collection techniques". This view was not shared by the Victorian Statistician, however, who "could not conceive that under any well devised system of census collection so large a proportion of the population could be overlooked".

Pell [44] [45] reported the results in 1867 and in 1879 of mortality studies undertaken in New South Wales for the periods 1856 to 1861 and 1860 to 1875 respectively. In 1867, he comments on the observed rounding of ages in census returns previously discussed, and states that "totals for periods of five years….may be relied upon as sufficiently near the truth". He also says that there "is no doubt that most of the deaths have been registered, and the corresponding ages stated more accurately than the censuses". In 1879, he says "There is every reason to believe that the returns of births and deaths are as accurate as can be expected in any case, and that the results of the censuses are as trustworthy as is usually the case". He also concludes that for census counts of children under five years of age, "There seems a wide-spread and unaccountable propensity to return young children as a year older than they really are". This conclusion was based on reconciling birth and death records, "which are most accurately kept", but his analysis ignored the effects of both immigration and emigration. The maximum magnitude of the derived undercount (and overcount) error rates for this age group was approximately 8%.

In 1884, Burridge [46] produced a table of rates of mortality for the period 1870 to 1881 derived from data from a number of Australian colonies. He excluded South Australian data from his calculations because it was not available with sufficiently detailed age distributions. This technical difficulty is addressed in Section 2.13 by using a statistical analysis procedure not available to Burridge. For children under five years of age, he also states, as did Pell, that "the census returns for this period cannot be trusted", but unlike Pell he does not provide a quantification of the error rate. He makes no other comment on the quality of the data that he used.

The weak and strong evidence presented in this section for the reliability of census and mortality data has been interpreted for this thesis as a range of values for δ_0 (undercount in deaths), for δ_2 (variability in population size) and for δ_3 (undercount in population size). These values are shown in Table 2.5.2, and the 24 possible combinations are contained in <<bootstrap.txt>> (*included* CD-rom).

Table 2.5.2: Values selected for δ_0, δ_2 and δ_3

$$\begin{matrix} \delta_0 & & \delta_2 & & \delta_3 \\ \begin{Bmatrix} 1.00 \\ 1.05 \\ 1.10 \end{Bmatrix} & \otimes & \begin{Bmatrix} 0 \\ .02 \end{Bmatrix} & \otimes & \begin{Bmatrix} 1 \\ 1.02 \\ 1.05 \\ 1.10 \end{Bmatrix} \end{matrix}$$

A particular combination might be considered the most appropriate choice for the years around a specific census. For example, {1, .02, 1.02} would seem to be a reasonable choice for the data from the years 1990 to 1992, while {1.05, .02, 1.05} is probably a fairer representation for the data from the years 1870 to 1872. The value of 1.10 for both δ_0 and δ_3 is included to represent what is considered the worst possible extreme, although commentaries do not really suggest that the data collection processes have underestimated true values to the extent of 10%.

The actuarial practice of data smoothing, or graduation, over adjacent age values to reduce any effect of the “psychological peculiarity” of excessive population and death counts has not been adopted for this thesis. This feature of the data, when it occurs, is subsumed into the estimate of the standard error of the estimate of current expected life.

2.6 Current life table analyses for the period 1971-1996

The presentation of the current life table analyses of the data compiled for this thesis is in the reverse chronological order of the years of the censuses on which the current life table analyses are focused. Following the overview presented in Section 2.5, it is assumed that the most recent data is more accurate and reliable than that from the earlier part of the 20th century, which in turn, is more accurate and reliable than that from the colonial era. Census years are grouped on the basis of common features of the associated data sets.

The sources of the data for this census grouping are shown in Table 2.6.1.

Table 2.6.1: Sources of population and mortality data for the period 1971-1996

Census Year	Population	Deaths
1971	ABS Census 1971	1970 : SRSA1975-76 (ages 0-5), SRSA1970-71 (ages 5+) 1971 : SRSA1975-76 (ages 0-5), SRSA1971-72 (ages 5+) 1972 : SRSA1975-76 (ages 0-5), SRSA1972-73 (ages 5+)
1976	ABS Census 1976	1975 : SRSA1975-76 1976 : ABS private 1977 : ABS private
1981	ABS Census 1981	1980 : ABS private 1981 : ABS private 1982 : ABS private
1986	ABS commissioned tabulation	1985 : ABS Deaths SA1985* 1986 : ABS Deaths SA1986* 1987 : ABS Deaths SA1987*
1991	ABS commissioned tabulation	1990 : ABS Demography SA1990* 1991 : ABS Demography SA1991* 1992 : ABS Demography SA1992*
1996	ABS commissioned tabulation	1995 : ABS Demography SA1995* 1996 : ABS Demography SA1996* 1997 : ABS Demography SA1997*

* Augmented by privately commissioned ABS tabulation for ages over 95 years

For this census grouping, the distributions of census population counts and the distributions of the number of deaths are tabulated at every single-year age level from birth to 105 years for each gender. This is the required form for the calculation of complete current life tables. Simple accumulation over specified five-year age ranges produces data in the appropriate form for the calculation of abridged current life tables. This census grouping is also characterized by the availability of official estimates of current expected life at birth for South Australia, calculated by either the ABS or by the Australian Government Actuary (AGA), or, in some instances, by both.

Each of the twelve data files (*see* Section 2.3 and *included* CD-rom) for this census grouping has been analyzed by the methods of Section 2.4, using the bootstrap specifications discussed in Section 2.5, producing two output files (*see* Section 2.4 and *included* CD-rom) for each data file.

A complete listing of << CELQM1996.txt >> is presented in Table 2.6.2 as a typical example of the output files produced by << current.f >>. The structure of these files is

- an introductory section that identifies the analysis, and gives the total population size and the total number of deaths determined from the data file.

- a further two major sections, presenting bootstrap results from a complete life table analysis (*i.e.* single-year age data) and from an abridged life table analysis (*i.e*. a mixture of single-year and grouped-years age data) (*see* Section 1.7).

- within each major section, bootstrap results are given for each set of values for $\delta_0, \delta_2, \delta_3$ and B specified by << bootstrap.txt >>. The two columns headed "Expected Life" and "SE" give the corresponding bootstrap arithmetic mean and bootstrap standard deviation. The three columns headed "Percentile Points" are the 5^{th} percentile, the 50^{th} percentile (median) and the 95^{th} percentile estimated from the empirical bootstrap distribution of B values. (*see* Section 1.11.3; bootstrap algorithm, step 5)

- within each major section, the first line of results includes a value enclosed in [ ] under the column headed "Expected Life". This is the current expected life calculated independently from the bootstrap estimation procedure, and is included for validation of the bootstrap computational process.

Table 2.6.2: A typical output file: << CELQM1996.txt >>

CELQM1996

Current Expected Life for Males centred on the census of 1996

The total size of the population is 698799
The total number of deaths is 5989

Complete Life Table
Bootstrap Estimation

Delta Values			B	Expected Life	SE	Percentile Points		
Delta0	Delta2	Delta3				5%	50%	95%
1.00	0.00	1.00	5001	[75.33] 75.33	0.17	75.05	75.33	75.62
1.05	0.00	1.00	5001	74.77	0.18	74.48	74.77	75.06
1.10	0.00	1.00	5001	74.45	0.18	74.15	74.44	74.75
1.00	0.02	1.00	5001	75.32	0.17	75.04	75.32	75.61
1.05	0.02	1.00	5001	74.77	0.17	74.50	74.77	75.06
1.10	0.02	1.00	5001	74.45	0.18	74.15	74.44	74.75
1.00	0.00	1.02	5001	75.55	0.17	75.28	75.56	75.84
1.05	0.00	1.02	5001	75.00	0.18	74.71	75.00	75.30
1.10	0.00	1.02	5001	74.68	0.18	74.38	74.67	74.97
1.00	0.02	1.02	5001	75.55	0.17	75.28	75.55	75.84
1.05	0.02	1.02	5001	75.00	0.17	74.72	75.00	75.28
1.10	0.02	1.02	5001	74.68	0.18	74.39	74.67	74.98
1.00	0.00	1.05	5001	75.88	0.17	75.60	75.88	76.17
1.05	0.00	1.05	5001	75.33	0.17	75.05	75.32	75.61
1.10	0.00	1.05	5001	75.01	0.18	74.72	75.00	75.31
1.00	0.02	1.05	5001	75.89	0.17	75.61	75.89	76.17
1.05	0.02	1.05	5001	75.33	0.17	75.05	75.33	75.62
1.10	0.02	1.05	5001	75.00	0.18	74.72	75.00	75.31
1.00	0.00	1.10	5001	76.41	0.17	76.14	76.41	76.69
1.05	0.00	1.10	5001	75.86	0.17	75.58	75.86	76.15
1.10	0.00	1.10	5001	75.54	0.18	75.25	75.53	75.83
1.00	0.02	1.10	5001	76.42	0.17	76.15	76.42	76.68
1.05	0.02	1.10	5001	75.86	0.17	75.58	75.85	76.13
1.10	0.02	1.10	5001	75.54	0.18	75.26	75.53	75.84

Table 2.6.2 (continued)

Abridged Life Table
Bootstrap Estimation

Delta Values			B	Expected Life	SE	Percentile Points		
Delta0	Delta2	Delta3				5%	50%	95%
1.00	0.00	1.00	5001	[75.34] 75.35	0.17	75.08	75.35	75.64
1.05	0.00	1.00	5001	74.91	0.19	74.62	74.91	75.23
1.10	0.00	1.00	5001	74.79	0.19	74.48	74.78	75.11
1.00	0.02	1.00	5001	75.35	0.17	75.07	75.34	75.63
1.05	0.02	1.00	5001	74.92	0.19	74.63	74.91	75.23
1.10	0.02	1.00	5001	74.78	0.19	74.48	74.78	75.11
1.00	0.00	1.02	5001	75.57	0.17	75.29	75.57	75.85
1.05	0.00	1.02	5001	75.15	0.19	74.85	75.14	75.46
1.10	0.00	1.02	5001	75.02	0.19	74.71	75.01	75.35
1.00	0.02	1.02	5001	75.58	0.17	75.29	75.58	75.86
1.05	0.02	1.02	5001	75.15	0.19	74.84	75.14	75.46
1.10	0.02	1.02	5001	75.02	0.19	74.71	75.01	75.34
1.00	0.00	1.05	5001	75.91	0.18	75.62	75.91	76.20
1.05	0.00	1.05	5001	75.48	0.18	75.19	75.48	75.79
1.10	0.00	1.05	5001	75.36	0.19	75.05	75.35	75.67
1.00	0.02	1.05	5001	75.91	0.17	75.62	75.91	76.19
1.05	0.02	1.05	5001	75.48	0.18	75.19	75.47	75.79
1.10	0.02	1.05	5001	75.36	0.19	75.05	75.35	75.68
1.00	0.00	1.10	5001	76.44	0.17	76.17	76.44	76.72
1.05	0.00	1.10	5001	76.02	0.18	75.73	76.02	76.33
1.10	0.00	1.10	5001	75.90	0.19	75.59	75.89	76.21
1.00	0.02	1.10	5001	76.44	0.17	76.17	76.44	76.72
1.05	0.02	1.10	5001	76.02	0.18	75.73	76.01	76.33
1.10	0.02	1.10	5001	75.90	0.19	75.59	75.89	76.21

Bootstrap arithmetic means and standard errors have been extracted from twelve of the output files (6 census years x 2 genders: selecting from the complete life table and the q-method, with bootstrap specification $\delta = (1,0,1)$). These values, with standard errors in parentheses, are displayed for the combinations of census year and gender within the columns headed “Thesis” in Table 2.6.3.

Table 2.6.3: Estimates of current expected life for the period 1971-1996

Year	Male Thesis	Male ABS	Female Thesis	Female ABS
1971	68.67 (0.20)	69.42	75.58 (0.20)	75.58
1975		70.04		77.12
1976	70.06 (0.19)	70.27	76.93 (0.18)	77.24
1977		70.96		77.53
1978		70.87		78.58
1979		71.49		78.47
1980		71.75		78.79
1981	71.49 (0.19)	72.18	78.92 (0.18)	79.48
1983		72.41		79.47
1985		73.14		79.08
1986	72.98 (0.18)	73.45	79.48 (0.18)	79.81
1987		73.47		80.04
1988		73.59		80.27
1989		73.73		79.72
1990		74.05		80.21
1991	74.08 (0.17)	74.65	80.30 (0.17)	80.40
1992		75.05		80.92
1993		74.99		80.53
1994		75.11		81.33
1993-95		75.10		81.01
1996	75.33 (0.17)		81.34 (0.16)	
1994-96		75.30		81.34
1995-97		75.70		81.52
1996-98		76.02		81.64
1997-99		76.43		82.08

These estimates are equivalent to the conventionally quoted figures for current expected life. The standard errors summarise the effect of sampling variation in the observed number of deaths, conditional on the observed census population counts, on the estimates of current expected life. For both males and females, these standard errors decrease with increasing census year and with the corresponding increase in population size. This is a commonly observed statistical phenomenon. Expressed as approximate 95% confidence intervals, (*i.e.* ± 2 SE), the statistical accuracy of the estimated expectations ranges from .8 years in 1971 to .6 years in 1996, approximately.

Statistical comparisons are made between the male and female estimates of current expected life for each census year by using an asymptotic normally distributed z-score [47]. This statistic is defined as the difference between the two estimates divided by the square root of the sum of the square of the standard error of each estimate. Values less than -2 or greater than +2 are considered statistically significantly different at the (approximate) 5% level. The relevant values of the z-score are shown in Table 2.6.4 (Thesis: Male *vs* Female), clearly indicating that females have a consistently larger current expected life than males over the years included in this census grouping period.

Table 2.6.4: Z-score comparison of estimates of current expected life for the period 1971-1996

Census Year	Thesis Male *vs* Female	Male Thesis *vs* ABS	Female Thesis *vs* ABS
1971	-24.4	-2.7	0
1976	-26.2	-0.8	-1.2
1981	-28.4	-2.6	-2.2
1986	-25.5	-1.8	-1.3
1991	-25.9	-2.4	-0.4
1996	-25.7	0.1	0

The ABS has also published estimates of current expected life for this period in *Deaths, South Australia 1969-1989* and *Demography, South Australia, 1990-continuing*. These estimates are included in Table 2.6.3 for comparison with the thesis results. There are, however, important methodological differences to be considered in making these comparisons. For example, the ABS estimates for 1971 are based on the number of deaths registered in 1971 only, and on population sizes projected retrospectively from the census of 1981. Similarly, the ABS estimates for the census year 1986 are based on prospective population projections from the census of 1981, whereas the results for 1987 are based on a limited projection of population sizes from the census of 1986. In all instances prior to 1995, the number of deaths registered in a single year only, and not the average of the number of deaths registered over three consecutive years, have been used in the calculation of current expected life by the ABS. As a consequence of a joint project with the Australian Government Actuary [48], the ABS has been calculating life expectations since 1995 based on the number of deaths from three consecutive years. For example, the ABS estimates shown in Table 2.6.3 for the years 1994-96 are based on the number of deaths that occurred in 1994, 1995 and 1996, and on population sizes for 1996. The thesis estimates for 1996 are based on the number of deaths from 1995, 1996 and 1997, and on the census count for 1996. In this one situation, where the data components of the calculations are almost the same, the estimates of current expected life obtained for the thesis and by the ABS are practically identical.

Formal comparisons between thesis and ABS estimates are shown in Table 2.6.4 using z-score statistics (Male:Thesis *vs* ABS and Female:Thesis *vs* ABS). For the purposes of this comparison, the error in the ABS estimate is assumed to be equal to the corresponding thesis estimate of error. While some of these estimates of expected value are nominally statistically significantly different, considering the procedural differences discussed above, inspection and comparison of the results given in Table 2.6.3 and Table 2.6.4 provides further reasonable evidence of the validation of the results produced by the thesis computer program.

Estimates of current expected life calculated using the bootstrap specification $\delta = (1,0,1)$ have been extracted from the output files to compare differences resulting from life table type (complete *vs* abridged) and method of analysis (q-method *vs* u-method). These expectations, with standard errors in parentheses, are displayed for the various combinations of census year and gender in Table 2.6.5.

Table 2.6.5: Methodological comparisons for the period 1971-1996

Census Year	Life Table Type	Male		Female	
		q-method	u-method	q-method	u-method
1971	complete	68.67 (0.20)	68.67 (0.20)	75.58 (0.20)	75.60 (0.20)
	abridged	68.70 (0.20)	68.81 (0.20)	75.61 (0.20)	75.75 (0.20)
1976	complete	70.06 (0.19)	70.07 (0.19)	76.93 (0.18)	76.93 (0.19)
	abridged	70.08 (0.19)	70.19 (0.20)	76.96 (0.19)	77.10 (0.19)
1981	complete	71.49 (0.19)	71.49 (0.19)	78.92 (0.18)	78.93 (0.18)
	abridged	71.51 (0.19)	71.63 (0.19)	78.97 (0.18)	79.11 (0.18)
1986	complete	72.98 (0.18)	72.98 (0.18)	79.48 (0.18)	79.49 (0.18)
	abridged	73.00 (0.18)	73.13 (0.19)	79.51 (0.18)	79.64 (0.18)
1991	complete	74.08 (0.17)	74.08 (0.17)	80.30 (0.17)	80.31 (0.17)
	abridged	74.10 (0.18)	74.23 (0.18)	80.32 (0.17)	80.46 (0.17)
1996	complete	75.33 (0.17)	75.33 (0.17)	81.34 (0.16)	81.34 (0.16)
	abridged	75.35 (0.17)	75.47 (0.18)	81.37 (0.16)	81.50 (0.16)

A consistent pattern can be seen in the sub-table of estimates within each census year and gender combination, which can be expressed as

$$\text{“Expectation(c,q)} \leq \text{Expectation(a,q)} \approx \text{Expectation(c,u)} < \text{Expectation(a,u)”}$$

where “c” and “a” indicate results derived from a complete life table and an abridged life table respectively, and q and u have been previously defined. Differences between the first three types of estimate, when they occur, are numerically small and have a maximum difference of .05 years over all estimates shown in Table 2.6.5. The fourth type of estimate produces values that are consistently .15 years (approximately) larger than the other three. However z-scores (not shown), comparing the first type of estimate to each of the other three estimates, are all between zero and one. The four standard errors that are obtained from the combination of life table type and method of analysis are essentially equal, for each census year and gender.

The extensive number of bootstrap analyses that were undertaken to investigate the effect that under-reporting in the census population counts and number of deaths might have on the estimate of current expected life are summarised in Table 2.6.6.

Table 2.6.6: Robustness comparisons for the period 1971-1996

Census Year	Male			
	Standard	Population+2%	Deaths+10%	Population+10%
1971	68.67 (.20)	68.91 (.20)	67.68 (.21)	69.87 (.19)
1976	70.06 (.19)	70.30 (.19)	69.08 (.21)	71.24 (.19)
1981	71.49 (.19)	71.73 (.19)	70.53 (.20)	72.64 (.18)
1986	72.98 (.18)	73.21 (.18)	72.05 (.19)	74.10 (.17)
1991	74.08 (.17)	74.30 (.17)	73.18 (.18)	75.17 (.17)
1996	75.33 (.17)	75.55 (.17)	74.45 (.18)	76.41 (.17)

Census Year	Female			
	Standard	Population+2%	Deaths+10%	Population+10%
1971	75.58 (.20)	75.81 (.19)	74.65 (.21)	76.66 (.19)
1976	76.93 (.18)	77.14 (.19)	76.01 (.20)	77.98 (.19)
1981	78.92 (.18)	79.13 (.18)	78.06 (.19)	79.92 (.17)
1986	79.48 (.18)	79.70 (.18)	78.62 (.19)	80.50 (.17)
1991	80.30 (.17)	80.50 (.17)	79.48 (.18)	81.29 (.16)
1996	81.34 (.16)	81.54 (.16)	80.57 (.17)	82.30 (.15)

Four estimates of current expected life with standard error are shown for each census year and gender, calculated from complete life tables using the q-method. The columns in Table 2.6.6 are headed

- Standard: using bootstrap specification $\delta = (1,0,1)$. The estimates in this column are the standard or conventional estimates of current expected life previously discussed and shown in Table 2.6.4 and Table 2.6.5. They are repeated for convenience of comparison.

- Population+2%: using bootstrap specification $\delta = (1,0,1.02)$. As discussed in Section 2.5, post-enumeration surveys undertaken by the ABS suggest a “probable” census population undercount of about 2%. The observed number of deaths is assumed to be accurate.

- Deaths+10%: using bootstrap specification $\delta = (1.1,0,1)$. This specification examines one extreme situation, in which the observed number of deaths is assumed, for the purposes of this thesis, to be a gross undercount. The census population count is assumed to be accurate. The estimate of current expected life in these circumstances is consequently less than that calculated under the conditions of “Standard”.

- Population+10%: using bootstrap specification $\delta = (1,0,1.1)$. This specification examines the other extreme, in which the census population size is assumed to be a gross undercount. The observed number of deaths is assumed to be accurate. The estimate of current expected life in these circumstances is consequently greater than that calculated under the conditions of “Standard”.

By inspection of Table 2.6.6, it can be seen that

1) the effect of increasing the census population counts, in accordance with the ABS estimate of undercount, while using the observed number of deaths, is to increase the estimate of current expected life by (approximately) .2 years at each census, for both males and females. This increase is of the order of 1.5 standard errors of the estimate, or a relative increase of (approximately) .3%.

2) the effect of increasing the observed number of deaths by 10%, subject to the bounds specified in Section 1.11.5, while using the observed census population counts, is to reduce the estimates of current expected life by (approximately) 1 year at each census, or a maximum relative decrease of (approximately) 1.5% for males and (approximately) 1.3% for females.

3) the effect of increasing the observed census population counts by 10%, while using the observed number of deaths, is to increase the estimates of current expected life by (approximately) 1 year at each census, or a relative increase of (approximately) 1.5% for males and (approximately) 1.3% for females over the period.

It can also be seen from the relevant output files included on the CD-rom, that the standard errors of the estimates of current expected life are unchanged (to two decimal places) by imposing additional random variation on the census population counts. *i.e.* estimates of error resulting from using bootstrap specifications of the generic form $\delta = (x,.02,y)$.

2.7 Current life table analyses for 1961 and 1966

The degree of detail of the gender and age-specific population and mortality data for this census grouping is almost identical to the degree of detail of the data presented in Section 2.6, with one minor exception. The distributions of the census population counts are tabulated at single-year age levels from birth to 105 years, as are the distributions of the number of deaths with the exception of ages 100 to 105, which are tabulated as accumulated totals only. For each gender, the number of deaths for each single-year within this age range has been approximated by assuming an equal spread over these five years and using the average value of this group. Thus, taking the strictest interpretation of the available data, an abridged current life table is more consistent with the reported data than is a complete current life. (However there are only slight numerical differences between estimates using the q-method with each type of life table.) There are no known published official estimates of current expected life for South Australia for this period. The sources of the data for this census grouping are shown in Table 2.7.1.

Table 2.7.1: Sources of population and mortality data for 1961 and 1966

Census Year	Population	Deaths
1961	ABS Census 1961	1960 : SRSA1965-66 (ages 0-5), SRSA1960-61 (ages 5+) 1961 : SRSA1965-66 (ages 0-5), SRSA1961-62 (ages 5+) 1962 : SRSA1965-66 (ages 0-5), SRSA1962-63 (ages 5+)
1966	ABS Census 1966	1965 : SRSA1965-66 1966 : SRSA1975-76 (ages 0-5), SRSA1966-67 (ages 5+) 1967 : SRSA1975-76 (ages 0-5), SRSA1967-68 (ages 5+)

The pattern of analyses that was established in Section 2.6 is repeated in this and following sections. The four data files and eight output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.7.2 (*cf* Table 2.6.3 and Table 2.6.4) shows estimates of current expected life with standard error for each census year and gender, calculated from complete life tables using the q-method. For each census year, z-scores are used to compare the estimates of current expected life between genders. The estimate of current expected life for males is statistically significantly less than the estimate for females.

Table 2.7.2: Estimates of current expected life for 1961 and 1966

Census Year	Male	Female	Z-score
1961	68.62 (.22)	74.91 (.21)	-20.7
1966	68.57 (.21)	75.22 (.20)	-22.9

Table 2.7.3 (*cf* Table 2.6.5) displays the estimate of current expected life with standard error for each combination of life table type and method of analysis, for each census year and gender. The general relationship observed between estimates in Section 2.6 and expressed there as

"Expectation(c,q) $\leq$ Expectation(a,q) $\approx$ Expectation(c,u) < Expectation(a,u)"

is again evident in Table 2.7.3 although numerical differences between estimates are small and are statistically insignificant when standard errors are considered. There is no evidence of any appreciable difference in the size of the four standard errors that are obtained from the combination of life table type and method of analysis, for each census year and gender.

Table 2.7.3: Methodological comparisons for 1961 and 1966

Census Year	Life Table Type	Male		Female	
		q-method	u-method	q-method	u-method
1961	**complete**	68.62 (.22)	68.62 (.23)	74.91 (.21)	74.92 (.21)
	abridged	68.64 (.23)	68.76 (.22)	74.93 (.22)	75.07 (.22)
1966	**complete**	68.57 (.21)	68.59 (.20)	75.22 (.20)	75.22 (.20)
	abridged	68.59 (.21)	68.71 (.21)	75.23 (.21)	75.36 (.21)

Table 2.7.4(*cf* Table 2.6.6) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life.

Table 2.7.4: Robustness comparisons for 1961 and 1966

Census Year	Male			
	Standard	**Population+2%**	**Deaths+10%**	**Population+10%**
1961	68.62 (.22)	68.88 (.22)	67.54 (.24)	69.86 (.21)
1966	68.57 (.21)	68.83 (.21)	67.56 (.22)	69.80 (.20)

Census Year	Female			
	Standard	**Population+2%**	**Deaths+10%**	**Population+10%**
1961	74.91 (.21)	75.14 (.21)	73.95 (.22)	76.01 (.20)
1966	75.22 (.20)	75.45 (.20)	74.26 (.22)	76.31 (.20)

From Table 2.7.4, it can be seen that adjusting the census population counts by an amount estimated by the ABS to be an appropriate level of undercount results in an increase in the estimates of current expected life of (approximately) .2 years, for both census years and genders. An increase of 10% in the number of deaths produces a decrease in the estimates of current expected life of (approximately) 1 year, while an increase of 10% in the population counts produces an increase in the estimates of current expected life of (approximately) 1 year. These are average relative changes of (approximately) 1.4% over the period.

Inspection of the relevant output files included on the CD-rom shows that the standard errors of the estimates of current expected life are essentially unchanged by imposing additional random variation on the census population counts.

2.8 Current life table analyses for the period 1933-1954

The intended schedule of holding a population census every seven years was disrupted in this period by the advent of World War II, and consequently a census of the population was not taken in 1940. The economic consequences of the Great Depression 1929-1939 also ended the previous practice of producing a separate current life table for each State in favour of a combined national current life table only.

The distributions of census population counts for the three censuses included in this grouping are tabulated at single-year age levels from birth to 105 years. This level of detail is not available for the distributions of the number of deaths, which are tabulated at single-year age levels from birth until 15 years, and subsequently as accumulated five year totals for age levels [15-20), [20-25),…,[100-105], for the years 1946-48 and 1953-55. For the years 1932-34, the distribution of the number of deaths is tabulated at single-year age levels from birth until 5 years, and subsequently as accumulated five year totals for age levels [5-10), [10-15),..,[100-105]. After suitable aggregation of most of the census population counts and some of the mortality data, the reported data has been reorganised into an appropriate form for an abridged current life table analysis. The reported mortality data has been used to approximate the data required for a complete current life table analysis, by using the average value calculated from each given five year age grouping as the single-age value of the number of deaths for each year within that group. The sources of the data for this census grouping are shown in Table 2.8.1.

Table 2.8.1: Sources of population and mortality data for the period 1933-1954

Census Year	Population	Deaths
1933	ABS Census 1933	1932 : SRSA1937-38 Decennial returns 1928-37 1933 : SRSA1937-38 Decennial returns 1928-37 1934 : SRSA1937-38 Decennial returns 1928-37
1947	ABS Census 1947	1946 : SRSA1955-56 (ages 0-5,15+), ABS Aust Demography No 64 (ages 5-15) 1947 : SRSA1955-56 (ages 0-5,15+), ABS Aust Demography No 65 (ages 5-15) 1948 : SRSA1955-56 (ages 0-5,15+), ABS Aust Demography No 66 (ages 5-15)
1954	ABS Census 1954	1953 : SRSA1955-56 (ages 0-5,15+), ABS Aust Demography No 71 (ages 5-15) 1954 : SRSA1955-56 (ages 0-5,15+), ABS Aust Demography No 72 (ages 5-15) 1955 : SRSA1955-56 (ages 0-5,15+), ABS Aust Demography No 73 (ages 5-15)

The six data files and twelve output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.8.2 (*cf* Table 2.6.3 and Table 2.6.4) shows estimates of current expected life with standard error for each census year and gender, calculated using the approximated data discussed above for complete life tables using the q-method. For each census year, z-scores are used to compare the estimates of current expected life between genders. The estimate of current expected life for males is statistically significantly less than the estimate for females. Although official estimates of current expected life are not available for this period, P.C. Wickens [49] analysed mortality for the six Australian states, and the expectations that he calculated for South Australia are also included in Table 2.8.2. Z-scores are used to compare his results with those obtained for this thesis, using error assumptions discussed in Section 2.6. Wickens presumably used data very similar to that obtained for this thesis, and he applied various sophisticated actuarial smoothing techniques as part of his calculations. The comparable estimates of current expected life obtained for this thesis and by Wickens are very close numerically and not statistically significantly different.

Table 2.8.2: Estimates of current expected life for the period 1933-1954

Census Year	Male	Female	Z-score
1933	65.34 (.34)	68.04 (.35)	-5.5
1947	67.17 (.27)	71.28 (.26)	-11.0
1954	67.72 (.25)	73.07 (.24)	-15.4
Wickens 1953-55	67.82		-.3
		73.09	-.1

Table 2.8.3 (*cf* Table 2.6.5) shows estimates of current expected life with standard error for each combination of life table type and method of analysis, for each census year and gender. The general relationship observed between estimates in previous sections and expressed there as

“Expectation(c,q) $\leq$ Expectation(a,q) $\approx$ Expectation(c,u) < Expectation(a,u)”

is again evident in Table 2.8.3 although numerical differences between estimates are small and are statistically insignificant when standard errors are considered. There is no evidence of any appreciable difference in the size of the four standard errors that are obtained from the combination of life table type and method of analysis, for each census year and gender. It is also worthwhile noting that the differences between estimates calculated from complete current life tables and abridged current life tables, using the q-method, are no larger in Table 2.6.5 than they are in Table 2.8.3. In Section 2.6, the number of deaths is tabulated at a known single-year age level, whereas in this section the tabulation of the number of deaths at the single-year age level is achieved by substituting group averages as approximations in some age groups.

Table 2.8.3: Methodological comparisons for the period 1933-1954

Census Year	Life Table Type	Male		Female	
		q-method	u-method	q-method	u-method
1933	**complete**	65.34 (.34)	65.34 (.34)	68.04 (.35)	68.06 (.35)
	abridged	65.37 (.34)	65.49 (.35)	68.09 (.36)	68.22 (.36)
1947	**complete**	67.17 (.27)	67.18 (.27)	71.28 (.26)	71.29 (.27)
	abridged	67.19 (.27)	67.31 (.27)	71.32 (.27)	71.45 (.28)
1954	**complete**	67.72 (.25)	67.72 (.25)	73.07 (.24)	73.08 (.24)
	abridged	67.74 (.25)	67.86 (.25)	73.11 (.25)	73.24 (.24)

Table 2.8.4(*cf* Table 2.6.6) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life.

From Table 2.8.4, it can be seen that adjusting the census population counts by an amount estimated by the ABS to be an appropriate level of undercount results in an increase in the estimates of current expected life of (approximately) .3 years, for all census years and genders. An increase of 10% in the number of deaths produces a decrease in the estimates of current expected life of (approximately) 1.2 years, while an increase of 10% in the population counts produces an increase in the estimates of current expected life of (approximately) 1.3 year. These are average relative changes of (approximately) 1.6% over the period.

Table 2.8.4: Robustness comparisons for the period 1933-1954

Census Year	**Male**			
	Standard	**Population+2%**	**Deaths+10%**	**Population+10%**
1933	65.34 (.34)	65.64 (.34)	63.96 (.36)	66.78 (.33)
1947	67.17 (.27)	67.44 (.27)	65.99 (.28)	68.46 (.26)
1954	67.72 (.25)	67.99 (.24)	66.60 (.26)	69.00 (.24)

Census Year	**Female**			
	Standard	**Population+2%**	**Deaths+10%**	**Population+10%**
1933	68.04 (.35)	68.35 (.36)	66.67 (.37)	69.47 (.34)
1947	71.28 (.26)	71.55 (.27)	70.13 (.28)	72.52 (.26)
1954	73.07 (.24)	73.31 (.24)	72.02 (.25)	74.21 (.23)

Inspection of the relevant output files included on the CD-rom shows that the standard errors of the estimates of current expected life are essentially unchanged by imposing additional random variation on the census population counts.

2.9 Current life table analyses for 1911 and 1921

The two censuses included in this period were the first censuses conducted by the Australian Bureau of Census and Statistics (established in 1905 and renamed the ABS in 1974) following the Federation of the Australian colonies in 1901.

For this census grouping, the distribution of the census population counts and the distributions of the number of deaths are tabulated at every single-year age level from birth to 105 years. This is the appropriate form for the direct calculation of complete current life tables, and simple accumulation over specified five-year age ranges produces data in the appropriate form for the calculation of abridged current life tables. The sources of the data for this census grouping are shown in Table 2.9.1.

Table 2.9.1: Sources of population and mortality data for 1911 and 1921

Census Year	Population	Deaths
1911	ABS Census 1911	1910 : ABS Aust Demography No 25 1911 : ABS Aust Demography No 29 1912 : ABS Aust Demography No 30
1921	ABS Census 1921	1920 : ABS Aust Demography No 38 1921 : ABS Aust Demography No 39 1922 : ABS Aust Demography No 40

The four data files and eight output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.9.2 (*cf* Table 2.6.3 and Table 2.6.4) shows estimates of current expected life with standard error for each census year and gender, calculated from complete life tables using the q-method. For each census year, z-scores are used to compare the estimates of current expected life between genders. The estimate of current expected life for males is statistically significantly less than the estimate for females.

Table 2.9.2: Estimates of current expected life for 1911 and 1921

Census Year	Male	Female	Z-score
ABS 1901-10	56.76		3.5
		60.39	3.4
1911	58.81 (.41)	62.41 (.42)	-6.1
1921	60.01 (.38)	63.53 (.39)	-6.5

Prior to the Australian life tables for 1921, estimates of current expected life had been produced using an earlier version of the now-standard methodology presented in this thesis. As commented by the Australian Statistician, R. Wilson, in the *Year Book of the Commonwealth Of Australia* 1936 "…it was resolved to base the investigation [*i.e.* new Australian life tables] upon the census of 1933, and the deaths in the calendar years 1932-34. This, of course, involves to a certain extent a break of continuity in the Australian series, for the tables prior to 1921 were based upon two censuses and the deaths of the intervening ten years. But tables based on the latter foundation suffer from the defect, which is ineradicable, that at the time of publication the mortality experience may be as much as ten years out of date." Some estimates produced by the ABS using this methodology for South Australian data are also shown in Table 2.9.2, designated ABS 1901-10, and the z-scores indicate that these estimates are statistically significantly lower than the corresponding thesis estimates for each gender for 1911. The relative underestimation by the obsolete procedure is (approximately) 3.5%, and estimates produced in this way have been included in review articles on Australian mortality [50]. In the *Statistician's report on the Census of 1921*, it states that "Tables for the triennium [*i.e.* 1920-1921] relating to the several States are in the course of preparation and the results will be published as the opportunity offers"; but these have not been forthcoming.

Table 2.9.3 (*cf* Table 2.6.5) displays the estimate of current expected life with standard error for each combination of life table type and method of analysis, for each census year and gender. The general relationship observed between estimates in previous sections and expressed there as

"Expectation(c,q) $\leq$ Expectation(a,q) $\approx$ Expectation(c,u) < Expectation(a,u)" is again evident in Table 2.9.3 although numerical differences are small and are statistically insignificant when standard errors are considered. There is no evidence of any appreciable difference in the size of the four standard errors that are obtained from the combination of life table type and method of analysis, for each census year and gender.

Table 2.9.3: Methodological comparisons for 1911 and 1921

Census Year	Life Table Type	Male		Female	
		q-method	u-method	q-method	u-method
1911	complete	58.81 (.41)	58.83 (.40)	62.41 (.42)	62.41 (.43)
	abridged	58.83 (.41)	58.94 (.40)	62.43 (.44)	62.54 (.43)
1921	complete	60.01 (.38)	60.01 (.38)	63.53 (.39)	63.53 (.39)
	abridged	60.03 (.38)	60.13 (.38)	63.55 (.39)	63.66 (.40)

Table 2.9.4 (*cf* Table 2.6.6) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life.

Table 2.9.4: Robustness comparisons for 1911 and 1921

Census Year	Male			
	Standard	Population+2%	Deaths+10%	Population+10%
1911	58.81 (.41)	59.18 (.41)	57.29 (.42)	60.58 (.40)
1921	60.01 (.38)	60.37 (.37)	58.51 (.39)	61.70 (.36)

Census Year	Female			
	Standard	Population+2%	Deaths+10%	Population+10%
1911	62.41 (.42)	62.76 (.43)	60.89 (.43)	64.10 (.42)
1921	63.53 (.39)	63.87 (.39)	62.00 (.40)	65.19 (.38)

From Table 2.9.4, it can be seen that adjusting the census population counts by an amount estimated by the ABS to be an appropriate level of undercount results in an increase in the estimates of current expected life of (approximately) .4 years, for all census years and genders. An increase of 10% in the number of deaths produces a decrease in the estimates of current expected life of (approximately) 1.5 years, while an increase of 10% in the population counts produces an increase in the estimates of current expected life of (approximately) 1.7 years. These are average relative changes of (approximately) 2.5% over the period.

Inspection of the relevant output files included on the CD-rom shows that the standard errors of the estimates of current expected life are essentially unchanged by imposing additional random variation on the census population counts.

2.10 Current life table analyses for the period 1876-1901

The distributions of the census population counts of the four censuses include in this grouping are tabulated at single-year age levels from birth to 105 years. This level of detail is not available for the distributions of the number of deaths, which are tabulated at single-year age levels from birth until five years, and subsequently as five-year totals for the age groups [5-10), [10-15),…,[100-105]. Aggregation into five-year age groups of most of the reported census population counts produces data in an appropriate form for an abridged current life table analysis. Conversely, the reported mortality data has been used to approximate the data required for a complete current life table analysis, by using the average value calculated from each given five year age group as an estimate of the single-age value of the number of deaths for each year within that group. The sources of the data for this census grouping are shown in Table 2.10.1.

Table 2.10.1: Sources of population and mortality data for the period 1876-1901

Census Year	Population	Deaths
1876	Census of Sth Aust SAPP 1877, No 73	1875 : SAGG 1876 1876 : SAGG 1877 1877 : SAGG 1878
1881	Census of Sth Aust SAPP 1881, No 74	1880 : SRSA 1880 1881 : SRSA 1881 1882 : SRSA 1882
1891	Census of Sth Aust SAPP 1891, No 74	1890 : SRSA 1890 1891 : SRSA 1891 1892 : SRSA 1892
1901	Census of Sth Aust SAPP 1901, No 74	1900 : SRSA 1900 1901 : SRSA 1901 1902 : SRSA 1902

The eight data files and sixteen output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.10.2 (*cf* Table 2.6.3) shows estimates of current expected life with standard error for each census year and gender, calculated from abridged life tables using the q-method. This is a change in practice from Section 2.6 to Section 2.9, where tabulated data is generally immediately compatible for use in complete life tables. It is considered that, for this and subsequent census groupings, abridged life tables are preferable to complete life tables because of the extent of approximation needed to adjust the reported number of deaths to values appropriate to single-year age levels. Table 2.10.2 also shows ABS estimates derived using the obsolete methodology discussed in Section 2.9, from mortality data over ten year periods.

Table 2.10.2: Estimates of current expected life for the period 1876-1901

	Male		Female	
Year	**Thesis**	**ABS**	**Thesis**	**ABS**
1876	43.57 (.58)		46.09 (.64)	
1881	47.62 (.50)		50.76 (.62)	
1881-90		50.61		53.81
1891	52.52 (.47)		54.98 (.52)	
1891-1900		53.02		56.10
1901	54.56 (.47)		57.72 (.50)	
1901-10		56.76		60.39

Table 2.10.3 (*cf* Table 2.6.4) shows, for each census year, z-score comparisons of the thesis estimates of current expected life for each gender. Z-score comparisons of the thesis and ABS estimates for each gender are also shown. For each census year in this period, the estimate of current expected life for males is statistically significantly less than the estimate for females. Also, as anticipated,

- the thesis estimates for 1881 are statistically significantly less than the corresponding ABS estimates for 1881-90.
- the thesis estimates for 1891 are between the corresponding ABS estimates for 1881-90 and 1891-1900.
- the thesis estimates for 1901 are between, and statistically significantly different from, the corresponding ABS estimates for 1891-1900 and 1901-10.

Table 2.10.3: Z-score comparison of estimates of current expected life for the period 1876-1901

Census Year	Thesis Male *vs* Female	Male Thesis *vs* ABS	Female Thesis *vs* ABS
1876	-2.9		
1881	-3.9	-4.2 (cf 1881-90)	-3.5
1891	-3.5	2.9 (cf 1881-90) -.8 (cf 1891-00)	1.6 1.5
1901	-4.6	2.3 (cf 1891-00) -3.3 (cf 1901-10)	2.3 -3.8

Table 2.10.4 (*cf* Table 2.6.5) shows estimates of current expected life with standard error for each combination of life table type and method of analysis, for each census year and gender.

Table 2.10.4: Methodological comparisons for the period 1876-1901

Census Year	Life Table Type	Male		Female	
		q-method	**u-method**	**q-method**	**u-method**
1876	**complete**	43.11 (.56)	43.16 (.56)	45.54 (.62)	45.57 (.61)
	abridged	43.57 (.58)	43.65 (.57)	46.09 (.64)	46.15 (.64)
1881	**complete**	47.24 (.51)	47.28 (.50)	50.18 (.58)	50.20 (.57)
	abridged	47.62 (.50)	47.73 (.50)	50.76 (.62)	50.88 (.62)
1891	**complete**	52.26 (.47)	52.26 (.47)	54.68 (.51)	54.68 (.51)
	abridged	52.52 (.47)	52.61 (.47)	54.98 (.52)	55.07 (.51)
1901	**complete**	54.37 (.48)	54.39 (.47)	57.49 (.50)	57.50 (.49)
	abridged	54.56 (.47)	54.64 (.47)	57.72 (.50)	57.81 (.49)

The general relationship observed between estimates in previous sections and expressed there as

“Expectation(c,q) $\leq$ Expectation(a,q) $\approx$ Expectation(c,u) < Expectation(a,u)”

is now

“Expectation(c,q) $\approx$ Expectation(c,u) < Expectation(a,q) < Expectation(a,u)”.

However, numerical differences are small and are statistically insignificant when standard errors are considered. The change in the relationship between estimates obtained under different conditions illustrates the effect of the extensive approximations that have been made to transform grouped five-year data into a form suitable for complete life table analysis. There is no evidence of any appreciable difference in the size of the four standard errors that are obtained from the combination of life table type and method of analysis, for each census year and gender.

Table 2.10.5 (*cf* Table 2.6.6) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life. However, in contrast to previous tables of this appearance in which the robustness of estimates was examined, definitions used for Table 2.10.5 and subsequent tables of this appearance are

- Standard: Estimates of current expected life are calculated from abridged life tables using the q-method and with bootstrap specification $\delta = (1,0,1)$. (In previous tables of this appearance, estimates are calculated from complete life tables.)
- Population+5%: Estimates of current expected life are calculated from abridged life tables using the q-method and with bootstrap specification $\delta = (1,0,1.05)$. (In previous tables of this appearance, estimates are calculated from complete life tables using the q-method and with $\delta = (1,0,1.02)$)
- Deaths+10%: Estimates of current expected life are calculated from abridged life tables using the q-method and with bootstrap specification $\delta = (1.1,0,1)$. (In previous tables of this appearance, estimates are calculated from complete life tables.)
- Population+10%: Estimates of current expected life are calculated from abridged life tables using the q-method and with bootstrap specification $\delta = (1,0,1.1)$. (In previous tables of this appearance, estimates are calculated from complete life tables.)

The preference for estimates obtained from abridged life tables rather than for those obtained from complete life tables is justified in the discussion relating to Table 2.10.2. The overview of the quality of colonial era data presented in Section 2.5 suggests that an undercount of 5% in census population counts for the 19th century is perhaps a more realistic figure than the undercount of 2% in census population counts assumed for the 20th century.

Table 2.10.5: Robustness comparisons for the period 1876-1901

Census Year	Male			
	Standard	**Population+5%**	**Deaths+10%**	**Population+10%**
1876	43.57 (.58)	44.82 (.57)	41.79 (.61)	46.01 (.57)
1881	47.62 (.50)	48.81 (.50)	46.01 (.54)	49.94 (.50)
1891	52.52 (.47)	53.55 (.46)	51.00 (.52)	54.55 (.46)
1901	54.56 (.47)	55.60 (.47)	53.07 (.52)	56.58 (.46)

Census Year	Female			
	Standard	**Population+5%**	**Deaths+10%**	**Population+10%**
1876	46.09 (.64)	47.32 (.63)	44.19 (.67)	48.51 (.63)
1881	50.76 (.62)	52.03 (.62)	48.92 (.64)	53.18 (.63)
1891	54.98 (.52)	56.02 (.51)	53.30 (.55)	57.02 (.50)
1901	57.72 (.50)	58.74 (.50)	56.12 (.54)	59.69 (.49)

From Table 2.10.5, it can be seen that adjusting the census population counts by the amount of 5%, assessed to be a reasonably appropriate level of undercount, results in an increase in the estimates of current expected life of (approximately) 1 year, for all census years and genders. An increase of 10% in the number of deaths produces a decrease in the estimates of current expected life of (approximately) 1.7 years, while an increase of 10% in the population counts produces an increase in the estimates of current expected life of (approximately) 2.2 years. These are average relative changes of (approximately) 4% over the period.

Inspection of the relevant output files included on the CD-rom shows that the standard errors of the estimates of current expected life are essentially unchanged by imposing additional random variation on the census population counts.

2.11 Current life table analyses for 1871

The analysis for this census year is presented separately from other censuses because of difficulties created by coarse tabulations in the distributions of census population counts and in the distributions of the number of deaths for the years 1870-72. For unknown and unstated reasons, official publications of mortality figures for this period use the age groupings [0-2), [2-5), [5-10), [10-30), [30-50) and [50+). For this reason, Burridge [46] excluded South Australia from his composite study of Australian mortality in 1884. The distributions of the census population counts are more detailed. They are tabulated at single-year age levels from birth until 15 years, as five-year totals for the age groups [15-20), [20-25),…,[75-80), and as an accumulation (over a nominal twenty five years) for the age group [80+). The sources of the data for this census year are shown in Table 2.11.1.

Table 2.11.1: Sources of population and mortality data for 1871

Census Year	Population	Deaths
1871	Census of Sth Aust SAPP 1871, No 9	1870-72 : SRSA 1873 RBDM : Deaths ; Books 29, 37-50

The lack of detail for the higher ages in the distributions of the reported census population counts for 1871 has been resolved for this thesis by reference to the census of 1876, which is the census closest in time that has the required level of detail. For each gender, Table 2.11.2 shows the total population size and the population count in the age group [80+) for the censuses of 1876 and 1871. The count in the age group [80+) is a very small percentage of each population. The observed population counts for the age groups [80-85), [85-90), [90-95), [95-100) and [100-105] are also shown for the census of 1876, with the corresponding proportion that these counts represent of the total count for the age group [80+). This conditional distribution has been applied to the census of 1871 to produce estimated population sizes for this census for the age groups [80-85), [85-90), [90-95), [95-100) and [100-105]. As an example of this process, the estimated population size for males in the age group [80-85) is calculated proportionally as 142 x 114/182 = 89. Similar calculations produce estimates of population size for the other age groups. The results are shown in Table 2.11.2.

Table 2.11.2: Estimated population sizes for 1871: Ages 80+

	Male				Female			
	1876		1871		1876		1871	
Population	110410		95288		102734		90164	
Age 80+	182	(.165%)	142	(.149%)	174	(.169%)	106	(.118%)
Age group	Observed		Estimated		Observed		Estimated	
80-85	114	.626	89		132	.759	80	
85-90	54	.297	43		36	.207	22	
90-95	10	.055	7		6	.034	4	
95-100	4	.022	3		0	0	0	
100-105	0	0	0		0	0	0	

The lack of detail in the reported mortality distributions has been resolved for this thesis by an extensive examination of the original individual death certificates of the RBDM in the archives held by State Records of South Australia. These handwritten certificates are arranged in the chronological order in which they were received at the central Adelaide Registry, and are bound in manuscript “death Books”. The division between entries of deaths occurring in one calendar year and the next is not always distinct. It is common to find that some records of deaths that occurred in December of one year are intermingled with the records of deaths that occurred in January and February of the following year. The cursive script of the handwriting of the collective pool of district Registry clerks of those times is not always clearly readable. However, within the resources and practical time constraints of this thesis, a concerted and diligent attempt has been made to examine each of the death certificates for the years 1870, 1871 and 1872, and consequently tabulate a single-year age distribution of the number of deaths for each gender and year. The officially reported distributions of the number of deaths for 1870-72 are shown in Table 2.11.3 under the columns headed SRSA. Table 2.11.3 also shows the distributions derived from the RBDM at the single-year level, but combined for this table into the age groupings used for official reporting. For each gender and calendar year, the two distributions are compared using a Chi-square test of homogeneity [47]. There is no statistical evidence that any sampled distribution differs from the corresponding reported official distribution, at this age grouping level.

Table 2.11.3: Comparison of the number of deaths from SRSA and RBDM for 1870-72

	Male					
	1870		1871		1872	
Age Group	**SRSA**	**RBDM**	**SRSA**	**RBDM**	**SRSA**	**RBDM**
0-2	647	578	651	603	771	734
2-5	94	105	70	90	102	122
5-10	45	63	44	54	68	65
10-30	134	124	152	146	154	160
30-50	198	204	190	203	191	205
50+	255	281	245	274	306	330
Total	1373	1355	1352	1370	1592	1616
	$\chi_5^2 = 9.12$	p = .104	$\chi_5^2 = 7.41$	p = .192	$\chi_5^2 = 4.10$	p = .535

	Female					
	1870		1871		1872	
Age Group	**SRSA**	**RBDM**	**SRSA**	**RBDM**	**SRSA**	**RBDM**
0-2	607	566	475	475	632	620
2-5	77	86	82	89	116	110
5-10	48	50	36	57	53	77
10-30	131	135	126	128	152	149
30-50	142	148	140	153	144	143
50+	165	180	167	191	205	221
Total	1170	1165	1026	1093	1302	1320
	$\chi_5^2 = 2.80$	p = .731	$\chi_5^2 = 5.12$	p = .401	$\chi_5^2 = 5.22$	p = .390

For each gender and calendar year, the mortality distribution derived from the RBDM is considered for the purposes of this thesis as a "sample", and the officially reported number of deaths for each age group is assumed to be correct. The derived or sampled single-year values have been used to conditionally distribute official group totals to individual single-year values. The age group [0-2) for males in 1870 is used as an illustration of this process. The sampled RBDM data for this age group has 440 deaths in the age range [0-1) and 138 deaths in the range [1-2), for a total of 578 deaths. Thus the officially reported SRSA1873 total of 647 deaths for the group [0-2) is allocated proportionally into the age range [0-1) as 647 x 440/578 = 493, and into the age range [1-2) as 647 x 138/578 = 154. Similar allocation procedures are applied to the other five official age groupings, and the resulting 105 single-year values for the number of deaths are accumulated into the conventional age groupings for an abridged life table *i.e.* [0-1), [1-2),…,[4-5),[5-10), [10-15),…,[100-105]. To be consistent with previous sections, the average value of each of the five-year age groupings is used as the single-year value of the number of deaths within that group as an approximation for complete life table analysis. The distributions of the number of deaths sampled from the RBDM are shown in Table 2.11.4 using the age groupings for a conventional abridged life table analysis. This table also shows the estimated distributions of the number of deaths resulting from the allocation procedure.

The composite (census counts and estimated sizes) distributions of population size, and the estimated distributions of the number of deaths have been tabulated using the age groupings for a conventional abridged life table analysis to produce the two data files for this section. The data is internally consistent, in the sense that the average number of deaths over the three calendar years is less than the population size each age. (The one exception is for females aged [95-100), with a population size of 0 and an average number of deaths of 1.) The two data files and the four output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.11.4: Comparison of sampled and estimated number of deaths for 1870-72

Male	Number of deaths sampled from RBDM			Estimated number of deaths		
Age Group	1870	1871	1872	1870	1871	1872
0-1	440	464	521	493	501	547
1-2	138	139	213	154	150	224
2-3	56	38	62	50	30	52
3-4	27	34	36	24	26	30
4-5	22	18	24	20	14	20
5-10	63	54	65	45	44	68
10-15	30	34	47	32	36	45
15-20	23	32	28	25	33	27
20-25	36	28	37	39	29	36
25-30	35	52	48	38	54	46
30-35	43	55	45	42	51	42
35-40	46	53	49	45	51	46
40-45	67	55	55	64	51	51
45-50	48	40	56	47	37	52
50-55	62	58	52	55	52	48
55-60	48	59	61	44	53	57
60-65	48	45	64	44	40	59
65-70	47	43	54	43	38	50
70-75	41	35	44	36	31	41
75-80	16	20	25	15	18	23
80-85	10	7	14	9	6	13
85-90	5	3	13	5	3	12
90-95	0	3	2	0	3	2
95-100	4	1	1	4	1	1
100-105	0	0	0	0	0	0
Total	1355	1370	1616	1373	1352	1592

Table 2.11.4 (continued)

Female	Number of deaths sampled from RBDM			Estimated number of deaths		
Age Group	1870	1871	1872	1870	1871	1872
0-1	417	349	461	447	349	470
1-2	149	126	159	160	126	162
2-3	49	37	55	44	34	58
3-4	20	29	37	18	27	39
4-5	17	23	18	15	21	19
5-10	50	57	77	48	36	53
10-15	36	24	33	35	24	34
15-20	33	31	37	32	31	38
20-25	38	36	31	37	35	32
25-30	28	37	48	27	36	48
30-35	37	38	36	36	35	36
35-40	55	46	34	52	41	34
40-45	32	40	43	31	37	43
45-50	24	29	30	23	27	31
50-55	30	42	44	28	37	41
55-60	28	25	35	26	22	32
60-65	36	32	37	33	28	34
65-70	40	31	38	37	27	35
70-75	19	23	38	17	20	35
75-80	15	18	19	14	16	18
80-85	9	13	4	7	10	4
85-90	2	3	4	2	3	4
90-95	1	1	2	1	1	2
95-100	0	3	0	0	3	0
100-105	0	0	0	0	0	0
Total	1165	1093	1320	1170	1026	1302

Table 2.11.5 (*cf* Table 2.6.3 and Table 2.6.4) shows estimates of current expected life with standard error for each gender for 1871, calculated from abridged life tables using the q-method. A z-score is used to compare the estimates of current expected life between genders. The estimate of current expected life for males is statistically significantly less than the estimate for females.

Table 2.11.5: Estimates of current expected life for 1871

Census Year	Male	Female	Z-score
1871	48.69 (.67)	52.19 (.77)	-3.4
Burridge 1870-81	46.47		2.1
		49.64	2.3

Table 2.11.5 also shows estimates for the period 1870 to 1881 calculated by Burridge [46] from data obtained from the colonies of Victoria, New South Wales and Queensland. These estimates are statistically significantly lower than the corresponding thesis estimates, although it is difficult to conclude whether this is a real difference in mortality or an artifact of two different statistical methodologies. Burridge uses the records of births and deaths of children under the age of five years to construct population counts for this group, "since [he considers] the census returns for this period cannot be trusted", and modifies the death returns truncated for ages above 80 by the "corresponding Healthy English rate". The corresponding thesis counterpart to Burridge's method is the proportional allocation of 1871 census total counts using conditional distributions derived from the census of 1876. Burridge also uses the actuarial graphical method of graduation or smoothing of the data.

Table 2.11.6 (*cf* Table 2.6.5) shows estimates of current expected life with standard error for each combination of life table type and method of analysis, for each gender. The general relationship observed between estimates in the previous section and expressed there as

"Expectation(c,q) $\approx$ Expectation(c,u) < Expectation(a,q) < Expectation(a,u)"

is still evident. However, numerical differences are small and are statistically insignificant when standard errors are considered. There is no evidence of any appreciable difference in the size of the four standard errors that are obtained from the combination of life table type and method of analysis, for each gender.

Table 2.11.6: Methodological comparisons for 1871

Life Table Type	Male		Female	
	q-method	u-method	q-method	u-method
complete	48.63 (.65)	48.65 (.66)	52.10 (.70)	52.12 (.71)
abridged	48.69 (.67)	48.78 (.66)	52.19 (.77)	52.28 (.75)
Doering (A)	48.71 (.67)		52.45 (.77)	
Doering (B)	49.14 (.68)		52.85 (.79)	
Doering (C)	53.54 (.79)		57.78 (.91)	

Table 2.11.6 also includes three additional estimates for each gender, denoted Doering (A), (B) & (C). Doering [51] considers the construction of what he calls skeleton life tables that are based on a much smaller number of age groupings than is traditional for an abridged life table analysis. He employs advanced techniques for ages less than five years and for ages greater than seventy-five. He concludes that the number of age groupings can be as low as seven and from "a large experience indicates that the method can be recommended to health officers as likely to give sufficiently good results [*i.e.* expected life] for their purposes". While the specific detail of his method has not been used here, three estimates have been included to illustrate the effect of age groups used in a current life table analysis on the resulting estimate of current expected life. In Table 2.11.6, the estimate

- Doering (A) uses the age groupings [0-2), [2-5), [5-10), [10-30), [30-50), [50-55), [55-60),…,[100-105]. This set of age groupings is a composite of those used for the official reporting of deaths under the age of fifty, and of the conventional five-year groupings used for ages over fifty. Estimated population sizes and estimated numbers of deaths are used in some of these age groupings.

- Doering (B) uses the age groupings [0-2), [2-5), [5-10), [10-30), [30-50), [50-55), [55-60),…,[75-80), [80-100),[100-105]. Estimated numbers of deaths are required in only some of these age groupings and reported census population counts are used in all age groupings.

- Doering (C) uses the age groupings [0-2), [2-5), [5-10), [10-30), [30-50), [50+). These are the age groupings of the officially reported death distributions. Only reported census population counts and reported number of deaths are required with these age groupings.

For each gender, the numerical difference between the abridged estimate and the Doering series of estimates becomes progressively larger, although the differences between abridged and Doering (A), and abridged and Doering (B), are not statistically significant using z-scores (not shown). The estimate Doering (C) is markedly larger than the corresponding abridged estimate, indicating the relative importance for detailed mortality distributions in the higher ages compared to the lower ages.

Table 2.11.7 (*cf* Table 2.10.5) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life.

Table 2.11.7: Robustness comparisons for 1871

	Standard	Population+5%	Deaths+10%	Population+10%
Male	48.69 (.67)	49.90 (.66)	46.89 (.67)	51.07 (.67)
Female	52.19 (.77)	53.37 (.77)	50.34 (.77)	54.49 (.79)

From Table 2.11.7, it can be seen that adjusting the census population counts by the amount of 5%, assessed to be a reasonably appropriate level of undercount, results in an increase in the estimate of current expected life of (approximately) 1.2 years, for each gender. An increase of 10% in the number of deaths produces a decrease in the estimates of current expected life of (approximately) 1.8 years, while an increase of 10% in the population counts produces an increase in the estimates of current expected life of (approximately) 2.3 years. These are average relative changes of (approximately) 3.5% and 4.5% respectively.

Inspection of the relevant output files included on the CD-rom shows that the standard errors of the estimates of current expected life are essentially unchanged by imposing additional random variation on the census population counts.

2.12 Current life table analyses for 1861 and 1866

The tabulations of the distributions of the census population counts of the censuses of 1861 and 1866, and some of the tabulations of the associated distributions of the number of deaths, have the unsatisfactory age grouping characteristics that are discussed in Section 2.11.

The census population counts for 1861 are tabulated using very unconventional age groupings; at single-year age levels from birth until 15 years; as a six-year total for the age group [15-21); at a single-year level for the age group [21-22); as five-year totals for the age groups [22-26), [26-31),…, [76-81); and as an accumulation (a nominal twenty four years) for the age group [81+). The census population counts for 1866 are tabulated at single-year age levels from birth until 15 years, as five-year totals for the age groups [15-20), [20-25),…, [75-80), and as accumulations (a nominal twenty five years) for the age group [80+).

For 1860, 1862 and 1865-67, the number of deaths is tabulated at single-year age levels from birth until five years, and subsequently as five-year totals for the age groups [5-10), [10-15),…, [100-105]. For 1861, the number of deaths is tabulated using the age groupings [0-2), [2-5), [5-10), [10-30), [30-50) and [50+).

The sources of data for this census grouping are shown in Table 2.12.1.

Table 2.12.1: Sources of population and mortality data for 1861 and 1866

Census Year	Population	Deaths
1861	Census of Sth Aust SAPP 1861, No 5	1860 : SAGG/RBDM 1861 1861 : RBDM : Deaths ; Books 9 & 13 SRSA 1864 1862 : SAGG/RBDM 1863
1866	Census of Sth Aust SAPP 1866-67, No 8	1865 : SAGG/RBDM 1866 1866 : SAGG/RBDM 1867 1867 : SAGG/RBDM 1868

For this thesis, the census population counts for 1861 have been proportionally redistributed into the conventional age groups for an abridged life table analysis by using the arithmetical calculations displayed in the following schematic:

Total for conventional age group…	Totals for reported age groups…
[15-20) =	$\frac{5}{6}$ [15-21)
[20-25) =	$\frac{1}{6}$ [15-21) + [21-22) + $\frac{3}{4}$ [22-26)
[25-30) =	$\frac{1}{4}$ [22-26) + $\frac{4}{5}$ [26-31)
[30-35) =	$\frac{1}{5}$ [26-31) + $\frac{4}{5}$ [31-36)
etc	etc
[75-80) =	$\frac{1}{5}$ [71-76) + $\frac{4}{5}$ [76-81)
[80+) =	$\frac{1}{5}$ [76-81) + [81+)

This re-allocation process assumes a uniform distribution of counts for single ages within each reported age group. The census population counts at single-year levels under the age of 15 remain unchanged.

For each gender and census year, Table 2.12.2 (*cf* Table 2.11.2) shows the allocation of the total for the reported age group [80+) into estimates for the age groups [80-85), [85-90), …, [100-105] using the conditioned observed age distribution from the census of 1876. The methodology is discussed in detail in Section 2.11.

Table 2.12.2: Estimated population sizes for 1861 and 1866: Ages 80+

	Male		**Female**	
	1861	**1866**	**1861**	**1866**
Population	64643	85626	61680	77975
Age 80+	38	74	29	61
Age group	**Estimated**	**Estimated**	**Estimated**	**Estimated**
80-85	24	46	22	46
85-90	11	22	6	13
90-95	2	4	1	2
95-100	1	2	0	0
100-105	0	0	0	0

The lack of detail in the reported age groupings of the mortality distributions for 1861 has been resolved by applying the principles of the sampling procedure discussed in detail in Section 2.11. For each gender, the distribution of the number of deaths for 1861 and published in SRSA 1864 is shown in Table 2.12.3 (*cf* Table 2.11.3). Data obtained from a sampling of the death records of the RBDM for 1861 is also shown, tabulated into the same age groupings as the official figures. The two distributions are compared using a Chi-square test of homogeneity. There is no statistical evidence that the reported and sampled distributions differ, at this age grouping level.

Table 2.12.3: Comparison of the number of deaths from SRSA and RBDM for 1861

Age Group	Male		Female	
	SRSA 1864	RBDM 1861	SRSA 1864	RBDM 1861
0-2	578	535	486	460
2-5	111	135	92	107
5-10	35	45	35	47
10-30	99	100	101	104
30-50	163	157	97	108
50+	100	109	54	65
Total	1086	1081	865	891
	$\chi_5^2 = 5.75$	$p = .331$	$\chi_5^2 = 4.87$	$p = .432$

For each gender, the distribution of the number of deaths sampled from the RBDM for 1861 is shown in Table 2.12.4 using the age groupings for a conventional abridged life table analysis. This table also shows the estimated distribution of the number of deaths resulting from the conditional allocation procedure, described in detail in Section 2.11, when applied to the age-group totals of SRSA 1864.

Table 2.12.4: Comparison of sampled and estimated number of deaths for 1861

	Male		Female	
Age Group	**Sampled**	**Estimated**	**Sampled**	**Estimated**
0-1	375	405	332	351
1-2	160	173	128	135
2-3	60	49	48	41
3-4	45	37	28	24
4-5	30	25	31	27
5-10	45	35	47	35
10-15	22	22	18	17
15-20	13	13	18	17
20-25	24	24	29	28
25-30	41	40	39	39
30-35	41	43	29	27
35-40	38	39	25	22
40-45	34	35	27	24
45-50	44	46	27	24
50-55	26	24	18	15
55-60	31	27	12	10
60-65	28	26	12	10
65-70	11	10	7	6
70-75	6	6	9	7
75-80	3	3	7	6
80-85	2	2	0	0
85-90	2	2	0	0
90-95	0	0	0	0
95-100	0	0	0	0
100-105	0	0	0	0
Total	1081	1086	891	865

The composite (*i.e.* census counts, age group reallocations, and estimated sizes) distributions of population size, and the reported and estimated distributions of the number of deaths have been tabulated using the age groupings for a conventional abridged life table analysis to produce the four data files for this section. The data is internally consistent, in the sense that the average number of deaths over the three calendar years encompassing a census year is less than the population size at each age. To be consistent with previous sections, the average value of each of the five-year age groupings is used as the single-year value within that group as an approximation for complete life table analysis, for both population and mortality distributions. The four data files and the eight output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.12.5 (*cf* Table 2.6.3 and Table 2.6.4) shows estimates of current expected life with standard error for each census year and gender, calculated from abridged life tables using the q-method. For each census year, z-scores are used to compare the estimates of current expected life between genders. The estimate of current expected life for males is statistically significantly less than the estimate for females. Table 2.12.5 also shows an estimate of expected life calculated by Pell [44] from census returns for the colony of New South Wales for the years 1856 to 1861. Pell does not differentiate between genders (he labels his result "persons") and his estimate is statistically significantly less than the thesis estimates of current expected life for both males and females for 1861. For z-score calculations, it is assumed that the error of the Pell estimate is equal to the error of the corresponding thesis estimate.

Table 2.12.5: Estimates of current expected life for 1861 and 1866

Census Year	Male	Female	Z-score
Pell 1856-61	45.58		-2.0
		45.58	-4.1
1861	48.55 (1.06)	51.88 (1.20)	-2.1
1866	47.39 (.86)	50.34 (.88)	-2.4

Table 2.12.6 (*cf* Table 2.6.5) shows estimates of current expected life with standard error for each combination of life table type and method of analysis, for each census year and gender.

Table 2.12.6: Methodological comparisons for 1861 and 1866

Census Year	Life Table Type	Male		Female	
		q-method	u-method	q-method	u-method
1861	**complete**	48.40 (0.82)	48.34 (0.83)	51.59 (0.84)	51.39 (0.83)
	abridged	48.55 (1.06)	48.63 (1.05)	51.88 (1.20)	51.71 (1.17)
1866	**complete**	47.27 (.73)	47.30 (.72)	50.20 (.76)	50.22 (.75)
	abridged	47.39 (.86)	47.47 (.86)	50.34 (.88)	50.43 (.89)

The general relationship observed between estimates in the previous section and expressed there as

“Expectation(c,q) $\approx$ Expectation(c,u) < Expectation(a,q) < Expectation(a,u)”

is still evident, with the exception of the results for females in 1861. However, numerical differences are small and are statistically insignificant when standard errors are considered. The effect of the strategy that has been used in this thesis when a value of ${}_j\hat{q}_k$ or $\hat{u}(k + j/2)$ is undefined can be seen in the estimates of error shown in Table 2.12.6. For each census year and gender, the estimate of error calculated from an abridged life table is appreciably larger than the corresponding error estimate calculated from a complete life table.

Table 2.12.7 (*cf* Table 2.10.5) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life.

Table 2.12.7: Robustness comparisons for 1861 and 1866

Census Year	Male			
	Standard	**Population+5%**	**Deaths+10%**	**Population+10%**
1861	48.55 (1.06)	49.99 (1.09)	46.42 (1.01)	51.34 (1.11)
1866	47.39 (0.86)	48.75 (0.87)	45.34 (0.83)	50.07 (0.89)

Census Year	Female			
	Standard	**Population+5%**	**Deaths+10%**	**Population+10%**
1861	51.88 (1.20)	53.30 (1.22)	49.72 (1.15)	54.52 (1.21)
1866	50.34 (0.88)	51.62 (0.89)	48.35 (0.87)	52.83 (0.90)

From Table 2.12.7, it can be seen that adjusting the census population counts by the amount of 5%, assessed to be a reasonably appropriate level of undercount, results in an increase in the estimates of current expected life of (approximately) 1.4 years, averaged over census year and gender. An increase of 10% in the number of deaths produces a decrease in the estimates of current expected life of (approximately) 1.6 years, while an increase of 10% in the population counts produces an increase in the estimates of current expected life of (approximately) 2.7 years. These are average relative changes of (approximately) 3.2% and 5.5% respectively.

Inspection of the relevant output files included on the CD-rom shows that the standard errors of the estimates of current expected life are essentially unchanged by imposing additional random variation on the census population counts.

2.13 Current life table analyses for 1851 and 1855

The tabulations of the distributions of the census population counts of the censuses of 1851 and 1855, and the tabulations of nearly all of the associated distributions of the number of deaths, have the unsatisfactory age grouping characteristics that are discussed in detail in Section 2.11.

The census population counts are tabulated using the very irregular age groupings of [0-2), [2-7), [7-14), [14-21), [21-45), [45-60) and [60+). According to Stevenson [36], these age groupings reflect "the social significance placed upon certain ages at the time: notably 14 representing the school leaving age and 21 the age of majority." As Stevenson also says "the awkward breakdown of age structure … substantially complicates and obscures comparison with population censuses after 1861."

For 1850-52 and 1854-55, the number of deaths is tabulated using the age groupings [0-2), [2-5), [5-10), [10-30), [30-50) and [50+). For 1856, the number of deaths is tabulated at single-year age levels from birth until five years, and subsequently as five-year totals for the age groups [5-10), [10-15),…, [100-105].

The sources of data for this census grouping are shown in Table 2.13.1.

Table 2.13.1: Sources of population and mortality data for 1851 and 1855

Census Year	Population	Deaths
1851	Census of Sth Aust SRSA 1856	1850-52 : SRSA 1859 RBDM : Deaths ; Books 1 & 2
1855	Census of Sth Aust SAPP 1860, No 5	1854-55 : SRSA 1861 RBDM : Deaths ; Book 3 1856 : SAGG/RBDM 1857

The overall lack of detail in the reported age groupings of the census population counts for this period has been resolved for this thesis in the following manner.

For ages between birth and two years:

For each gender, Table 2.13.2 shows the census population counts in the age groups [0-1) and [1-2) for the three censuses immediately following this period. A Chi-square test of homogeneity indicates that the proportion of counts in these two age groups is statistically indistinguishable between the censuses of 1861, 1866 and 1871. Proportions calculated from data combined over these three censuses have been applied to the census population count for the age group [0-2) to estimate population sizes for the age groups [0-1) and [1-2), for 1851 and 1855.

Table 2.13.2: Estimated population sizes for 1851 and 1855: Ages 0-2

Census year	Male			Female		
	Age group			Age group		
	0-1	1-2	0-2	0-1	1-2	0-2
1861	2525	2336	4861	2549	2380	4929
1866	3082	2939	6021	3119	2850	5969
1871	3189	3136	6325	3090	3006	6096
	$\chi_2^2 =$ 2.58	p=.275		$\chi_2^2 =$ 3.04	p=.219	
Combined	8796	8411	17207	8758	8236	16994
Proportion	.512	.488		.516	.484	
	Estimated		**Observed**	**Estimated**		**Observed**
1851	1101	1050	2151	1151	1080	2231
1855	1577	1504	3081	1667	1563	3230

For ages between two and sixty years:

For each census year and gender, the following procedure has been used to produce single-year estimates of population size for all ages between 2 and 60.

Table 2.13.3 shows the census population counts and corresponding value of the empirical cumulative distribution function for the reported age groups. A Burr Type-12 function [52], [53] specified as

$$F(x\ ;\ c, k) = 1 - (1 + x^{c})^{-k} \text{ where } x, c, k \geq 0$$

has been fitted to each empirical distribution function. Program AR from the BMDP Statistical Software package [54] was used to estimate numerical values for the parameters c and k by the method of least squares. *i.e.* by minimizing

$$\sum_{i=1}^{7} \{C(a_i) - F(a_i\ ;\ c, k)\}^2 \qquad a_i = 2, 7, 14, 21, 45, 60, 105$$

for c and k, where $C(a_i)$ is the empirical cumulative distribution function calculated at age a_i.

Table 2.13.3: Census population counts for 1851 and 1855: Ages 2-60

	Male				Female			
Age Group	**1851**		**1855**		**1851**		**1855**	
0-2	2151	.061	3081	.070	2231	.080	3230	.078
2-7	4734	.195	6689	.223	4592	.246	6682	.239
7-14	4778	.330	6984	.383	4472	.407	6798	.403
14-21	3847	.439	4903	.495	3945	.550	6024	.548
21-45	17106	.924	17855	.904	10707	.936	15592	.924
45-60	2293	.989	3492	.984	1526	.991	2620	.987
60+	400	1.000	712	1.000	257	1.000	524	1.000
Total	35309		43716		27730		41470	

Using **F** to denote a generic Burr function with estimated parameters, an appropriately conditioned form of **F** was applied separately to each of the age groups [2-7), [7-14), [14-21), [21-45) and [45-60) to produce single-year estimates of population size. Using a specific example outlines the procedure.

For the age group [2-7) from the census of 1851 for males, the census population count of 4734 has been proportioned into single-year estimates of population size through the arithmetical calculations indicated in the following schematic:

Estimated population size for age group…	Calculated by….
[2-3) =	4734 x (**F**(3)-**F**(2)) / (**F**(7)-**F**(2))
[3-4) =	4734 x (**F**(4)-**F**(3)) / (**F**(7)-**F**(2))
[4-5) =	4734 x (**F**(5)-**F**(4)) / (**F**(7)-**F**(2))
[5-6) =	4734 x (**F**(6)-**F**(5)) / (**F**(7)-**F**(2))
[6-7) =	4734 x (**F**(7)-**F**(6)) / (**F**(7)-**F**(2))

Similar operations, extending over the age group range and with **F** suitably conditioned by the age group boundaries, have been used for each of the other age groups [7-14), [14-21), [21-45) and [45-60), resulting in single-year estimates of population size for all ages between 2 and 60.

For ages greater than sixty years:
The procedure described for the age group [80+) in Section 2.11 has been adapted to estimate population sizes for the age groups [60-65), [65-70),…, [100-105] from the census population count for the age group [60+). For each gender, Table 2.13.4 (*cf* Table 2.11.2) shows the census population counts for the age groups [60-65), [65-70),…, [100-105] from the census of 1861, with the corresponding conditional proportions of the total of the age group [60+). These conditional proportions from 1861 have been applied to the census population count for the age group [60+) from the censuses of 1851 and 1855, to produce the estimated population sizes shown in Table 2.13.4.

Table 2.13.4: Estimated population sizes for 1851 and 1855: Ages 60+

	Male					
	1861		1855		1851	
Population	64640		43716		35309	
Age 60+	1409	(2.17%)	712	(1.63%)	400	(1.13%)
Age group	**Observed**		**Estimated**		**Estimated**	
60-65	733	.520	371		208	
65-70	362	.257	183		103	
70-75	190	.135	96		54	
75-80	85	.060	43		24	
80-85	24	.017	12		7	
85-90	12	.009	6		4	
90-95	2	.001	1		0	
95-100	1	.000	0		0	
100-105	0	0	0		0	

	Female					
	1861		1855		1851	
Population	61678		41470		27730	
Age 60+	1136	(1.84%)	524	(1.26%)	257	(0.93%)
Age group	**Observed**		**Estimated**		**Estimated**	
60-65	587	.517	271		133	
65-70	311	.274	144		70	
70-75	144	.127	67		33	
75-80	64	.056	29		14	
80-85	23	.020	10		5	
85-90	6	.005	3		2	
90-95	1	.000	0		0	
95-100	0	0	0		0	
100-105	0	0	0		0	

The lack of detail in the reported age groupings of the mortality distributions for 1850-52 and 1854-55 has been resolved by applying the principles of the sampling procedure discussed in detail in Section 2.11. For each gender, the distributions of the number of deaths published in SRSA 1859 and SRSA 1861 are shown in Table 2.13.5 and Table 2.13.6. (*cf* Table 2.11.3) Data obtained from a sampling of the death records of the RBDM for 1850-52 and 1854-55 is also shown, tabulated into the same age groupings as the official figures.

Table 2.13.5: Comparison of the number of deaths from SRSA and RBDM for 1850-52

Age Group	Male					
	1850		1851		1852	
	SRSA 1859	RBDM 1850	SRSA 1859	RBDM 1851	SRSA 1859	RBDM 1852
0-2	310	293	278	243	301	313
2-5	36	47	25	23	31	35
5-10	18	18	17	18	19	20
10-30	90	88	81	71	68	63
30-50	76	85	96	86	86	85
50+	32	33	35	35	34	38
Total	562	564	532	476	539	554
	$\chi_5^2 = 2.48$	p = .780	$\chi_5^2 = 0.56$	p = .990	$\chi_5^2 = 0.72$	p = .982

Age Group	Female					
	1850		1851		1852	
	SRSA 1859	RBDM 1850	SRSA 1859	RBDM 1851	SRSA 1859	RBDM 1852
0-2	244	235	261	232	256	258
2-5	37	37	27	28	37	36
5-10	16	23	10	10	17	17
10-30	43	61	58	56	54	47
30-50	57	51	59	62	70	69
50+	13	12	26	25	34	37
Total	410	419	441	413	468	464
	$\chi_5^2 = 4.82$	p = .438	$\chi_5^2 = 0.94$	p = .968	$\chi_5^2 = 0.62$	p = .987

Table 2.13.6: Comparison of the number of deaths from SRSA and RBDM for 1854-55

Age Group	Male			
	1854		1855	
	SRSA 1861	RBDM 1854	SRSA 1861	RBDM 1855
0-2	411	414	472	440
2-5	51	57	79	67
5-10	20	19	28	20
10-30	65	59	88	72
30-50	86	91	112	93
50+	64	65	61	45
Total	697	705	840	737
	$\chi_5^2 = 0.76$	p = .979	$\chi_5^2 = 2.50$	p = .776

Age Group	Female			
	1854		1855	
	SRSA 1861	RBDM 1854	SRSA 1861	RBDM 1855
0-2	357	343	413	352
2-5	61	67	61	49
5-10	22	22	18	15
10-30	74	71	108	104
30-50	74	75	107	96
50+	39	38	51	39
Total	627	616	758	655
	$\chi_5^2 = 0.55$	p = .990	$\chi_5^2 = 1.22$	p = .943

The corresponding distributions are compared using a Chi-square test of homogeneity. There is no statistical evidence that the reported and sampled distributions differ, at this level of age grouping.

For each gender, the distributions of the number of deaths sampled from the death records of the RBDM for 1850-52 and 1854-55 are shown in Table 2.13.7 and Table 2.13.8 using the age groupings for a conventional abridged life table analysis. Also shown are the estimated distributions of the number of deaths resulting from the conditional allocation procedure (*see* Section 2.11) applied to the age-group totals given in Table 2.13.6.

Table 2.13.7: Comparison of sampled and estimated number of deaths for 1850-52

Male	Number of deaths sampled from RBDM			Estimated number of deaths		
Age Group	1850	1851	1852	1850	1851	1852
0-1	216	171	229	229	196	220
1-2	77	72	84	81	82	81
2-3	22	9	19	17	10	17
3-4	14	11	9	11	12	8
4-5	11	3	7	8	3	6
5-10	18	18	20	18	17	19
10-15	10	10	6	10	11	6
15-20	17	8	9	17	9	10
20-25	33	31	17	34	36	18
25-30	28	22	31	29	25	34
30-35	23	28	24	21	31	24
35-40	20	26	15	18	29	15
40-45	24	18	23	21	20	23
45-50	18	14	23	16	16	24
50-55	15	15	13	14	15	12
55-60	4	5	5	4	5	4
60-65	5	8	8	5	8	7
65-70	2	1	1	2	1	1
70-75	3	3	6	3	3	5
75-80	2	2	3	2	2	3
80-85	2	1	1	2	1	1
85-90	0	0	1	0	0	1
90-95	0	0	0	0	0	0
95-100	0	0	0	0	0	0
100-105	0	0	0	0	0	0
Total	564	476	554	562	532	539

Table 2.13.7 (continued)

Female Age Group	Number of deaths sampled from RBDM 1850	1851	1852	Estimated number of deaths 1850	1851	1852
0-1	172	183	179	179	206	178
1-2	63	49	79	65	55	78
2-3	23	18	16	23	17	17
3-4	6	5	13	6	5	13
4-5	8	5	7	8	5	7
5-10	23	10	17	16	10	17
10-15	8	10	8	6	10	9
15-20	14	9	9	10	9	10
20-25	24	16	15	16	17	18
25-30	15	21	15	11	22	17
30-35	19	17	21	22	16	22
35-40	19	21	19	21	20	19
40-45	11	14	17	12	13	17
45-50	2	10	12	2	10	12
50-55	4	12	14	5	13	13
55-60	1	6	7	1	6	6
60-65	2	1	4	2	1	4
65-70	2	3	0	2	3	0
70-75	0	2	4	0	2	4
75-80	2	1	6	2	1	6
80-85	0	0	1	0	0	1
85-90	1	0	0	1	0	0
90-95	0	0	1	0	0	0
95-100	0	0	0	0	0	0
100-105	0	0	0	0	0	0
Total	419	413	464	410	441	468

Table 2.13.8: Comparison of sampled and estimated number of deaths for 1854-55

Male Age Group	Number of deaths sampled from RBDM 1854	 1855	Estimated number of deaths 1854	 1855
0-1	303	320	301	343
1-2	111	120	110	129
2-3	31	31	28	36
3-4	17	22	15	26
4-5	9	14	8	17
5-10	19	20	20	28
10-15	5	10	6	12
15-20	14	10	15	12
20-25	13	21	14	26
25-30	27	31	30	38
30-35	20	27	19	33
35-40	25	26	23	31
40-45	25	24	24	29
45-50	21	16	20	19
50-55	22	17	21	23
55-60	14	10	14	14
60-65	8	8	8	11
65-70	8	6	8	8
70-75	8	3	8	4
75-80	3	0	3	0
80-85	1	1	1	1
85-90	1	0	1	0
90-95	0	0	0	0
95-100	0	0	0	0
100-105	0	0	0	0
Total	705	737	697	840

Table 2.13.8 (continued)

Female Age Group	Number of deaths sampled from RBDM 1854	 1855	Estimated number of deaths 1854	 1855
0-1	237	237	247	278
1-2	106	115	110	135
2-3	35	27	32	34
3-4	21	14	19	17
4-5	11	8	10	10
5-10	22	15	22	18
10-15	10	8	10	8
15-20	10	12	11	12
20-25	29	44	30	46
25-30	22	40	23	42
30-35	25	35	24	39
35-40	22	26	22	29
40-45	15	21	15	23
45-50	13	14	13	16
50-55	11	11	12	14
55-60	5	10	5	13
60-65	9	6	9	8
65-70	5	10	5	13
70-75	4	0	4	1
75-80	2	1	2	1
80-85	1	1	1	1
85-90	1	0	1	0
90-95	0	0	0	0
95-100	0	0	0	0
100-105	0	0	0	0
Total	616	655	627	758

The estimated distributions of population size, and the reported and estimated distributions of the number of deaths have been tabulated using the age groupings for a conventional abridged life table analysis to produce the four data files for this section. The data is internally consistent, in the sense that the average number of deaths over the three calendar years encompassing a census year is less than the population size at each age. To be consistent with previous sections, the average value of each of the five-year age groupings is used as the single-year value within that group as an approximation for complete life table analysis, for both population and mortality distributions. The four data files and the eight output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.13.9 (*cf* Table 2.6.3 and Table 2.6.4) shows estimates of current expected life with standard error for each census year and gender, calculated from abridged life tables using the q-method. For each census year, z-scores are used to compare the estimates of current expected life between genders, and the estimate of current expected life for males is not statistically significantly different from the estimate for females.

Table 2.13.9 also shows the estimate of expected life calculated by Pell [44] for a "person" in the colony of New South Wales for the years 1856 to 1861. For 1855, Pell's estimate is not statistically significantly different from the thesis estimate for males, but it is statistically significantly different from the thesis estimate for females. For z-score calculations, it is assumed that the error of the Pell estimate is equal to the error of the corresponding thesis estimate.

Table 2.13.9: Estimates of current expected life for 1851 and 1855

Census Year	Male	Female	Z-score
1851	45.73 (1.24)	47.62 (1.31)	-1.1
1855	47.68 (1.27)	50.00 (1.27)	-1.3
Pell 1856-61	45.58		1.2
		45.58	2.5

Table 2.13.10 (*cf* Table 2.6.5) shows estimates of current expected life with standard error for each combination of life table type and method of analysis, for each census year and gender. The general relationship between the estimates of current expected life discussed in previous sections has been perturbed as a result of the extensive use of the two separate imputation processes required by the large number of undefined estimates of ${}_j\hat{q}_k$ and $\hat{u}(k + j/2)$ occurring at the higher ages of these two census years.

Table 2.13.10 also shows estimates designated as "Doering", arising from the concept of a skeleton life table that was introduced and discussed in Section 2.11. These estimates have been calculated from a tabulation of the population and mortality data using the reported age groupings of [0-2), [2-7), [7-14), [14-21), [21-45) and [45-60) of the census population counts, combined with standard five-year groupings of [60-65), [65-70), …[100-105]. For each census year and gender, the estimates from a conventional abridged current life table and from a "Doering" current life table are not statistically significantly different. (z-scores, not shown)

Table 2.13.10: Methodological comparisons for 1851 and 1855

Census Year	Life Table Type	Male		Female	
		q-method	u-method	q-method	u-method
1851	**complete**	45.58 (0.95)	45.58 (0.94)	47.45 (0.93)	47.49 (0.92)
	abridged	45.73 (1.24)	45.83 (1.22)	47.62 (1.31)	47.72 (1.29)
	Doering	45.76 (1.24)		47.80 (1.34)	
1855	**complete**	47.48 (0.92)	47.38 (0.89)	49.73 (0.89)	49.66 (0.87)
	abridged	47.68 (1.27)	47.60 (1.21)	50.00 (1.27)	49.89 (1.20)
	Doering	47.93 (1.28)		50.07 (1.24)	

Table 2.13.11 (*cf* Table 2.10.5) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life.

Table 2.13.11: Robustness comparisons for 1851 and 1855

Census Year	Male			
	Standard	**Population+5%**	**Deaths+10%**	**Population+10%**
1851	45.73 (1.24)	47.02 (1.25)	43.61 (1.20)	48.23 (1.22)
1855	47.68 (1.27)	49.09 (1.28)	45.49 (1.21)	50.41 (1.28)

Census Year	Female			
	Standard	**Population+5%**	**Deaths+10%**	**Population+10%**
1851	47.62 (1.31)	48.85 (1.35)	45.50 (1.24)	49.98 (1.35)
1855	50.00 (1.27)	51.31 (1.26)	47.83 (1.19)	52.46 (1.26)

From Table 2.13.11, it can be seen that adjusting the census population counts by the amount of 5%, assessed to be a reasonably appropriate level of undercount, results in an increase in the estimates of current expected life of (approximately) 1.3 years, averaged over census year and gender. An increase of 10% in the number of deaths produces a decrease in the estimates of current expected life of (approximately) 2.2 years, while an increase of 10% in the population counts produces an increase in the estimates of current expected life of (approximately) 2.5 years. These are average relative changes of (approximately) 4.6% and 5.2% respectively.

Inspection of the relevant output files included on the CD-rom shows that the standard errors of the estimates of current expected life are essentially unchanged by imposing additional random variation on the census population counts.

2.14 Current life table analyses for 1844 and 1846

The censuses of 1844 and 1846 were the first and second official censuses taken in the colony of South Australia. The tabulations of the distributions of the census population counts have the unsatisfactory age grouping characteristics that are discussed in detail in the previous sections, as have the tabulations of the associated distributions of the number of deaths. There are also some additional complications with these mortality distributions.

The census population counts for both years are tabulated using the age groupings of [0-2), [2-7), [7-14), [14-21), [21-45), [45-60) and [60+).

For each of the years 1844-47, the number of deaths is tabulated using the age groupings of "7 years & under", "7 to 14", "14 to 21", "21 to 30", "30 to 40", "40 to 50", "50 to 60", "60 to 70" and "70 to 83". The boundaries between age groupings are not clearly defined, having an apparent over-lap, and the tabulation is not differentiated by gender but is given for "persons" only. However, the total number of deaths is shown separately for each gender. For 1843, official age tabulations of the number of deaths have not been produced, and only the total number of deaths for each gender has been recorded. It is stated in *Statistical Returns for 1856*: "Note:- This Return merely shows the number of Births, Marriages and Deaths actually registered in the Province; there are, at present, no satisfactory data for estimating the number of those unregistered".

The sources of data for this census grouping are shown in Table 2.14.1.

Table 2.14.1: Sources of population and mortality data for 1844 and 1846

Census Year	Population	Deaths
1844	Census of Sth Aust SRSA 1856	1844-45 : SRSA 1847 RBDM : Deaths ; Book 1
1846	Census of Sth Aust SRSA 1856	1845-47 : SRSA 1847 RBDM : Deaths ; Book 1

The overall lack of detail in the reported age groupings of the census population counts for this period has been resolved for this thesis by applying the methodology established in Section 2.13. For each census year and gender, the broad age ranges "For ages between birth and two years", "For ages between two and sixty years" and "For ages over sixty years" have been analysed similarly to the corresponding age groups for 1851 and 1855. The essential features of this process are summarized in Table 2.14.2 (*cf* Table 2.13.2), Table 2.14.3 (*cf* Table 2.13.3) and Table 2.14.4 (*cf* Table 2.13.4), and the specific details are not repeated.

The lack of detail in the reported age groupings of the mortality distributions for 1843-45 and 1845-47 has been resolved by applying the principles of the sampling procedure discussed in detail in Section 2.11. The distributions of the number of deaths published in SRSA 1847 are shown in Table 2.14.5 and Table 2.14.6 (*cf* Table 2.11.3). Data obtained from a sampling of the death records of the RBDM for 1843-45 and 1845-47 is also shown, combined over genders and tabulated into the same age groupings as the official figures. The ambiguity in the official age group boundaries was clarified and the precise definition of the age groups is also indicated in these tables. The corresponding distributions are compared using a Chi-square test of homogeneity. There is no statistical evidence that the reported and sampled distributions differ, at this age grouping level. In the absence of official distributions of the number of deaths, only sampled distributions of the number of deaths for 1843 can be shown.

However, there appears to be a problem with both the reported and sampled data for 1843-45, in that the total number of deaths for both genders seems suspiciously low for these years when compared to the other early years of the colony. For each gender, Table 2.14.7 shows the total population size and total number of deaths for each census year in the period 1841 to 1876, with the gross percentage death rate included in parenthesis. The average gross percentage death rate calculated from the seven censuses from 1846 to 1876 is approximately 1.7% for males and 1.6% for females. These two values are considerably higher than the corresponding values based on the officially reported deaths for 1844-45, reinforcing the official warning about non-registration of deaths quoted previously in this section. A total number of deaths has therefore been estimated for each gender for 1843-45 by applying the average percentages to the appropriate total population size. These estimated totals are also shown in Table 2.14.7.

Table 2.14.2: Estimated population sizes for 1844 and 1846: Ages 0-2

Census year	Male			Female		
	Age group			Age group		
	0-1	1-2	0-2	0-1	1-2	0-2
1861	2525	2336	4861	2549	2380	4929
1866	3082	2939	6021	3119	2850	5969
1871	3189	3136	6325	3090	3006	6096
	$\chi_2^2 = 2.58$	p=.275		$\chi_2^2 = 3.04$	p=.219	
Combined	8796	8411	17207	8758	8236	16994
Proportion	.512	.488		.516	.484	
	Estimated		**Observed**	**Estimated**		**Observed**
1844	456	434	890	430	404	834
1846	522	497	1019	492	461	953

Table 2.14.3: Census population counts for 1844 and 1846: Ages 2-60

Age Group	Male				Female			
	1844		1846		1844		1846	
0-2	890	.093	1019	.080	834	.110	953	.099
2-7	1459	.247	2143	.250	1434	.298	2101	.316
7-14	1322	.385	1606	.376	1240	.461	1460	.468
14-21	922	.482	1088	.462	866	.575	981	.569
21-45	4431	.947	6111	.945	2935	.961	3696	.952
45-60	457	.995	629	.994	281	.998	410	.995
60+	44	1.000	74	1.000	18	1.000	49	1.000
Total	9525		12670		7608		9650	

Table 2.14.4: Estimated population sizes for 1844 and 1846: Ages 60+

	Male					
	1861		1846		1844	
Population	64640		12670		9525	
Age 60+	1409	(2.17%)	74	(0.58%)	44	(0.46%)
Age group	**Observed**		**Estimated**		**Estimated**	
60-65	733	.520	39		23	
65-70	362	.257	19		11	
70-75	190	.135	10		6	
75-80	85	.060	4		3	
80-85	24	.017	2		1	
85-90	12	.009	0		0	
90-95	2	.001	0		0	
95-100	1	.000	0		0	
100-105	0	0	0		0	

	Female					
	1861		1846		1844	
Population	61678		9650		7608	
Age 60+	1136	(1.84%)	49	(0.51%)	18	(0.24%)
Age group	**Observed**		**Estimated**		**Estimated**	
60-65	587	.517	25		9	
65-70	311	.274	13		5	
70-75	144	.127	6		2	
75-80	64	.056	3		1	
80-85	23	.020	1		1	
85-90	6	.005	1		0	
90-95	1	.000	0		0	
95-100	0	0	0		0	
100-105	0	0	0		0	

Table 2.14.5: Comparison of the number of deaths from SRSA and RBDM for 1843-45

Age grouping for...		Male and Female					
		1843		1844		1845	
SRSA	RBDM	SRSA	RBDM	SRSA	RBDM	SRSA	RBDM
7&under	[0-8)		80	81	81	147	148
7-14	[8-15)		7	4	4	8	8
14-21	[15-22)		6	4	5	6	6
21-30	[22-31)		17	17	15	19	14
30-40	[31-41)		24	20	20	28	27
40-50	[41-51)		10	6	6	22	23
50-60	[51-61)		6	5	5	3	3
60-70	[61-71)		3	2	2	3	3
70-83	[71-105]		0	1	1	2	2
	Total		153	140	139	238	234
				$\chi_4^2 =$ .11	p=.99	$\chi_5^2 =$ .75	.98

Table 2.14.6: Comparison of the number of deaths from SRSA and RBDM for 1845-47

Age grouping for...		Male and Female					
		1845		1846		1847	
SRSA	RBDM	SRSA	RBDM	SRSA	RBDM	SRSA	RBDM
7&under	[0-8)	147	148	244	237	317	310
7-14	[8-15)	8	8	15	15	20	16
14-21	[15-22)	6	6	6	6	16	19
21-30	[22-31)	19	14	19	18	39	38
30-40	[31-41)	28	27	31	34	53	54
40-50	[41-51)	22	23	26	27	32	28
50-60	[51-61)	3	3	12	11	12	13
60-70	[61-71)	3	3	5	6	4	4
70-83	[71-105]	2	2	2	2	2	2
	Total	238	234	360	356	495	484
		$\chi_5^2 =$ .75	p=.98	$\chi_5^2 =$ .26	p=.99	$\chi_6^2 =$ .97	p=.99

Table 2.14.7: Estimated total population sizes and total number of deaths for 1841-45

Year (census year in bold)	Male			
	Population		Deaths	
	SRSA	Estimated (interpolation)	SRSA	Estimated (.017Population)
1841	8195			141
1843		9082	* 80 [0.88]	155
1844	9526		75 [0.79]	164
1845		11098	143 [1.29]	200
1846	12670		208 [1.64]	
1847		17198	301 [1.75]	
1851	35309		532 [1.51]	
1855	43716		888 [2.03]	
1861	64643		1095 [1.69]	
1866	85626		1537 [1.80]	
1871	95288		1352 [1.42]	
1876	110410		1983 [1.80]	
			* RBDM	

Year (census year in bold)	Female			
	Population		Deaths	
	SRSA	Estimated (interpolation)	SRSA	Estimated (.016Population)
1841	6345			102
1843		7188	* 73 [1.02]	116
1844	7610		65 [0.85]	123
1845		8630	95 [1.10]	144
1846	9650		152 [1.58]	
1847		13266	194 [1.46]	
1851	27730		441 [1.59]	
1855	41469		775 [1.87]	
1861	61680		867 [1.41]	
1866	77975		1216 [1.56]	
1871	90165		1026 [1.14]	
1876	102734		1567 [1.53]	
			*RBDM	

For each gender and calendar year, the single-year age proportions determined from the distribution sampled from the RBDM have been applied to either the estimated or reported total number of deaths, as appropriate, to produce an estimated age distribution of the number of deaths. The distributions of the number of deaths sampled from the RBDM for 1843-45 and 1845-47, and the estimated distributions, are shown in Table 2.14.8 and Table 2.14.9 using the age groupings for a conventional abridged life table analysis.

The estimated distributions of population size, and the estimated distributions of the number of deaths, have been tabulated using the age groupings for a conventional abridged life table analysis to produce the four data files for this section. The data is internally consistent, in the sense that the average number of deaths over the three calendar years encompassing a census year is less than the population size at each age. To be consistent with previous sections, the average value of each of the five-year age groupings is used as the single-year value within that group as an approximation for complete life table analysis, for both population and mortality distributions.

The four data files and the eight output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.14.8: Comparison of sampled and estimated number of deaths for 1843-45

Male Age Group	Number of deaths sampled from RBDM			Estimated number of deaths		
	1843	1844	1845	1843	1844	1845
0-1	22	29	49	42	64	70
1-2	11	9	20	20	20	29
2-3	1	4	9	2	9	13
3-4	0	0	4	0	0	6
4-5	2	1	1	4	2	1
5-10	4	3	7	8	6	10
10-15	3	1	1	6	2	1
15-20	2	1	5	4	2	7
20-25	4	4	1	8	9	1
25-30	2	3	6	4	6	8
30-35	7	7	6	14	16	8
35-40	9	3	10	17	6	14
40-45	4	2	9	8	4	13
45-50	2	2	8	4	4	11
50-55	0	2	1	0	4	1
55-60	4	2	2	8	4	3
60-65	1	2	0	2	4	0
65-70	1	0	2	2	0	3
70-75	1	0	0	2	0	0
75-80	0	1	0	0	2	0
80-85	0	0	1	0	0	1
85-90	0	0	0	0	0	0
90-95	0	0	0	0	0	0
95-100	0	0	0	0	0	0
100-105	0	0	0	0	0	0
Total	80	76	142	155	164	200

Table 2.14.8 (continued)

Female	Number of deaths sampled from RBDM			Estimated number of deaths		
Age Group	1843	1844	1845	1843	1844	1845
0-1	29	16	34	45	32	52
1-2	6	14	20	9	27	30
2-3	5	3	3	8	6	5
3-4	0	1	0	0	2	0
4-5	1	0	1	2	0	2
5-10	0	1	5	0	2	7
10-15	3	3	1	5	6	2
15-20	3	1	0	5	2	0
20-25	3	4	2	5	8	3
25-30	5	2	4	8	4	6
30-35	2	7	5	3	14	8
35-40	6	4	3	10	8	5
40-45	6	3	3	10	6	5
45-50	1	0	5	2	0	8
50-55	1	1	3	2	2	5
55-60	1	1	1	2	2	2
60-65	0	1	0	0	2	0
65-70	1	1	1	0	2	2
70-75	0	0	0	0	0	0
75-80	0	0	1	0	0	2
80-85	0	0	0	0	0	0
85-90	0	0	0	0	0	0
90-95	0	0	0	0	0	0
95-100	0	0	0	0	0	0
100-105	0	0	0	0	0	0
Total	73	63	92	116	123	144

Table 2.14.9: Comparison of sampled and estimated number of deaths for 1845-47

Male	Number of deaths sampled from RBDM			Estimated number of deaths		
Age Group	1845	1846	1847	1845	1846	1847
0-1	49	80	109	70	80	112
1-2	20	33	35	29	33	37
2-3	9	2	10	13	2	10
3-4	4	6	7	6	6	7
4-5	1	2	6	1	2	6
5-10	7	11	16	10	11	16
10-15	1	4	9	1	4	9
15-20	5	2	8	7	2	8
20-25	1	7	10	1	7	11
25-30	6	6	12	8	6	12
30-35	6	9	16	8	9	16
35-40	10	14	19	14	14	19
40-45	9	9	12	13	9	12
45-50	8	8	7	11	8	7
50-55	1	2	8	1	2	8
55-60	2	6	6	3	6	6
60-65	0	3	1	0	3	1
65-70	2	2	1	3	2	1
70-75	0	1	1	0	1	1
75-80	0	1	0	0	1	0
80-85	1	0	2	1	0	2
85-90	0	0	0	0	0	0
90-95	0	0	0	0	0	0
95-100	0	0	0	0	0	0
100-105	0	0	0	0	0	0
Total	142	208	295	200	208	301

Table 2.14.9 (continued)

Female Age Group	Number of deaths sampled from RBDM 1845	1846	1847	Estimated number of deaths 1845	1846	1847
0-1	34	45	76	52	47	79
1-2	20	29	29	30	31	31
2-3	3	13	10	5	13	10
3-4	0	9	4	0	9	4
4-5	1	2	4	2	2	4
5-10	5	12	7	7	12	7
10-15	1	4	4	2	4	4
15-20	0	1	5	0	1	5
20-25	2	2	7	3	2	7
25-30	4	5	9	6	5	9
30-35	5	5	9	8	5	9
35-40	3	3	10	5	3	10
40-45	3	7	6	5	7	6
45-50	5	5	3	8	5	3
50-55	3	3	3	5	3	3
55-60	1	0	1	2	0	1
60-65	0	3	2	0	3	2
65-70	1	0	0	2	0	0
70-75	0	0	0	0	0	0
75-80	1	0	0	2	0	0
80-85	0	0	0	0	0	0
85-90	0	0	0	0	0	0
90-95	0	0	0	0	0	0
95-100	0	0	0	0	0	0
100-105	0	0	0	0	0	0
Total	92	148	189	144	152	194

Table 2.14.10 (*cf* Table 2.6.3 and Table 2.6.4) shows estimates of current expected life with standard error for each census year and gender, calculated from abridged life tables using the q-method. For each census year, z-scores are used to compare the estimates of current expected life between genders, and the estimate of current expected life for males is not statistically significantly different from the estimate for females.

Table 2.14.10 also shows estimates of the "mean after-lifetime" or expected value of the length of life at birth for 1841 given in the second English Life Table [55]. For each gender, the estimate for England for 1841 is not statistically significantly different from the thesis estimate for South Australia for 1844. For z-score calculations, it is assumed that the error of the English Life Table estimate is equal to the error of the corresponding thesis estimate.

Table 2.14.10: Estimates of current expected life for 1844 and 1846

Census Year	Male	Female	Z-score
ELT (1841)	40.0		.8
		42.0	1.5
1844	41.95 (1.56)	43.49 (1.60)	-.7
1846	41.57 (1.55)	46.67 (2.08)	-2.0

Table 2.14.11 (*cf* Table 2.6.5) shows estimates of current expected life with standard error for each combination of life table type and method of analysis, for each census year and gender. Table 2.14.11 also shows estimates designated as "Doering", discussed in detail in Section 2.11 and Section 2.13. These estimates have been calculated from a tabulation of the population and mortality data using the reported age groupings of [0-2), [2-7), [7-14), [14-21), [21-45) and [45-60) of the census population counts, combined with standard five-year groupings of [60-65), [65-70), …[100-105]. For each census year and gender, the estimates from a conventional abridged current life table and from a "Doering" current life table are not statistically significantly different (z-scores, not shown).

Table 2.14.11: Methodological comparisons for 1844 and 1846

Census Year	Life Table Type	Male		Female	
		q-method	u-method	q-method	u-method
1844	complete	41.81 (1.11)	41.81 (1.12)	43.34 (1.08)	43.38 (1.07)
	abridged	41.95 (1.56)	42.08 (1.57)	43.49 (1.60)	43.75 (1.61)
	Doering	42.20 (1.68)		43.80 (1.79)	
1846	complete	41.40 (1.13)	41.45 (1.14)	46.50 (1.22)	46.41 (1.21)
	abridged	41.57 (1.55)	41.70 (1.53)	46.67 (2.08)	46.76 (1.95)
	Doering	41.87 (1.62)		46.83 (2.18)	

Table 2.14.12 (*cf* Table 2.10.5) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life. The extent of estimation that has been undertaken for both population size and number of deaths may limit the interpretational value of this table. It has been included for completeness with other sections.

Table 2.14.12: Robustness comparisons for 1844 and 1846

Census Year	Male			
	Standard	Population+5%	Deaths+10%	Population+10%
1844	41.95 (1.56)	42.97 (1.59)	39.95 (1.52)	43.95 (1.59)
1846	41.57 (1.55)	42.70 (1.51)	39.41 (1.50)	43.83 (1.56)

Census Year	Female			
	Standard	Population+5%	Deaths+10%	Population+10%
1844	43.49 (1.60)	44.41 (1.56)	41.62 (1.54)	45.29 (1.59)
1846	46.67 (2.08)	47.95 (2.13)	44.27 (1.92)	49.17 (2.09)

Inspection of the relevant output files included on the CD-rom shows that the standard errors of the estimates of current expected life are essentially unchanged by imposing additional random variation on the census population counts.

2.15 Current life table analyses for 1841

The analysis undertaken in this section is speculative, since it is based on very limited population and mortality data. It is included both for its curiosity value, and as an attempt to complete the estimation of current expected life over the entire span of 170 years of European settlement in South Australia. The results presented here are not entirely inconsistent with the results shown in previous sections. The sources of the data used in this section are shown in Table 2.15.1.

Table 2.15.1: Sources of population and mortality data for 1841

Census Year	Population	Deaths
1841	House of Commons Parliamentary Paper No 505, V32, 1843.	1837-42 : Index of deaths, RBDM, 1959 1842 : RBDM, Deaths; Book 1

The population distributions shown in Table 2.15.2 are a consolidation of two "Official Statistical Returns relative to the progress of the colony of South Australia, at the termination of the year 1840" included in a *Despatch from Governor Grey to Lord John Russell* in October 1841. The separate tabulations within this return are labeled "Census of the Municipality of Adelaide" and "Census of the Country Districts". One country area has been excluded from Table 2.15.2 because only a total is given for each gender (Kangaroo Island, with 77 males and 13 females). For each gender, Table 2.15.2 shows the population count for each age group, and the associated percentage and cumulative proportion of the total count.

The population distributions shown in Table 2.15.2 have not been widely circulated, and are not reproduced within the SRSA series. While the official communiqué describes these figures as a "census", the returns of 1841 are not officially accorded this status. Stevenson [36] discusses the issues and details of the first attempts to enumerate the population of the colony, and describes the process producing the data given in Table 2.15.2 as "a primitive tabulation not unlike the early musters of New South Wales".

Table 2.15.2: Population counts from the "mustering" of 1841

Age group	Male			Female		
[0-7)	1651	20.1	.201	1593	25.1	.251
[7-14)	1083	13.2	.333	975	15.4	.405
[14-21)	822	10.0	.433	793	12.5	.530
[21-35)	3381	41.3	.846	2261	35.6	.886
[35-50)	1087	13.3	.979	624	9.8	.984
[50-105]	171	2.1	1.000	99	1.6	1.000
Total	8195			6345		

The overall lack of detail in the reported age groupings of the census population counts shown in Table 2.15.2 has been resolved for this thesis by applying the principles of the methodologies established in previous sections. Specifically,

For ages between birth and seven years:

For each gender, Table 2.15.3 (*cf* Table 2.11.2) shows, for the specified age groups, the estimated population size and corresponding conditional proportion of the population total of the age group [0-7) for 1844. These conditional proportions have been applied to the population total of the age group [0-7) for 1841 to produce the estimated population sizes for that year.

Table 2.15.3: Estimated population sizes for 1841: Ages 0-7

	Male			
	1844		1841	
Population	9525		8195	
Age [0-7)	2349	(24.7%)	1651	(20.1%)
Age group	Estimated		Estimated	
[0-1)	456	.194	320	
[1-2)	434	.185	305	
[2-3)	247	.105	173	
[3-4)	277	.118	195	
[4-5)	298	.127	210	
[5-7)	637	.271	448	

Table 2.15.3 (continued)

	Female			
	1844		1841	
Population	7608		6345	
Age [0-7)	2268	(29.8%)	1593	(25.1%)
Age group	**Estimated**		**Estimated**	
[0-1)	430	.190	303	
[1-2)	404	.178	284	
[2-3)	278	.123	196	
[3-4)	288	.127	202	
[4-5)	291	.128	204	
[5-7)	577	.254	404	

For ages between seven and fifty years:

For each gender, the procedures described in Section 2.13: "For ages between two and sixty years" were used to fit a Burr Type-12 function to the empirical cumulative distribution shown in Table 2.15.2. The appropriately conditioned form of the fitted function was used in conjunction with each of the population sizes for the age groups [7-14), [14-21), [21-35) and [35-50) to produce single-year estimates of population size for ages between seven and fifty years.

For ages greater than fifty years:
The procedure described for the age group [80+) in Section 2.11 has been adapted to estimate population sizes for the age groups [50-65), [55-60),…, [100-105] from the census population count for the age group [50+). For each gender, Table 2.15.4 (*cf* Table 2.11.2) shows the estimated population sizes for the age groups [50-55), [55-60),…, [100-105] from the census of 1844, with the corresponding conditional proportions of the total for the age group [50+). These conditional proportions from 1844 have been applied to the census population count for the age group [50+) from the censuses of 1841, to produce the estimated population sizes shown in Table 2.15.4.

Table 2.15.4: Estimated population sizes for 1841: Ages 50+

	Male			
	1844		1841	
Population	9525		8195	
Age 50+	298	(3.13%)	171	(2.08%)
Age group	**Estimated**		**Estimated**	
[50-55)	147	.493	85	
[55-60)	107	.359	61	
[60-65)	23	.077	13	
[65-70)	11	.037	6	
[70-75)	6	.020	3	
[75-80)	3	.010	2	
[80-85)	1	.003	1	
[85-90)	0	0	0	
[90-95)	0	0	0	
[95-100)	0	0	0	
[100-105]	0	0	0	

	Female			
	1844		1841	
Population	7608		6381	
Age 50+	176	(2.31%)	99	(1.55%)
Age group	**Estimated**		**Estimated**	
[50-55)	91	.517	50	
[55-60)	67	.381	38	
[60-65)	9	.051	5	
[65-70)	5	.028	3	
[70-75)	2	.011	1	
[75-80)	1	.006	1	
[80-85)	1	.006	1	
[85-90)	0	0	0	
[90-95)	0	0	0	
[95-100)	0	0	0	
[100-105]	0	0	0	

There is no officially reported mortality information for 1841 because civil registration of deaths was only introduced after March 1842. For each gender, Table 2.15.5 shows the mortality distributions tabulated from the extensive number of individual entries in the *Index to deaths: 1802-1842* produced by the RBDM in 1959 (*see* Section 2.1), and from a sampling of the earliest official death records of the RBDM.

Table 2.15.5: Sampled and estimated number of deaths for 1841

	Male					
		(Index)		**(RBDM)**	**Sample**	**Estimated**
Age group	**1837-39**	**1840**	**1841-42**	**part1842**	**1837-42**	**Deaths**
[0-1)	46	55	52	56	209	46
[1-2)	28	29	10	11	78	17
[2-3)	12	13	6	1	32	7
[3-4)	7	7	2	1	17	4
[4-5)	3	3	4	0	10	2
[5-10)	12	11	3	5	31	7
[10-15)	1	5	4	2	12	3
[15-20)	5	4	2	5	16	4
[20-25)	24	12	5	4	45	10
[25-30)	16	16	9	7	48	11
[30-35)	5	23	11	8	47	10
[35-40)	7	6	5	4	22	5
[40-45)	5	10	4	6	25	6
[45-50)	8	5	3	2	18	4
[50-55)	6	3	3	1	13	3
[55-60)	0	1	2	1	4	1
[60-65)	0	0	0	0	0	0
[65-70)	0	2	0	1	3	1
[70-75)	1	1	0	0	2	0
[75-80)	1	0	0	1	2	0
[80-85)	0	0	0	0	0	0
[85-90)	0	0	0	0	0	0
[90-95)	0	0	0	0	0	0
[95-100)	0	0	0	0	0	0
[100-105]	0	0	0	0	0	0
Total	187	206	125	116	634	141

Table 2.15.5 (continued)

Age group	Female (Index) 1837-39	1840	1841-42	(RBDM) part1842	Sample 1837-42	Estimated Deaths
[0-1)	41	45	40	37	163	35
[1-2)	20	21	12	14	67	15
[2-3)	10	12	5	2	29	6
[3-4)	5	6	4	0	15	3
[4-5)	2	4	0	0	6	1
[5-10)	6	10	4	2	22	5
[10-15)	2	3	5	1	11	2
[15-20)	5	6	1	0	12	3
[20-25)	10	7	3	1	21	5
[25-30)	11	12	12	5	40	9
[30-35)	16	11	9	6	42	9
[35-40)	4	4	3	4	15	3
[40-45)	2	2	3	4	11	2
[45-50)	1	3	0	0	4	1
[50-55)	0	3	0	1	4	1
[55-60)	2	1	0	0	3	1
[60-65)	0	0	0	0	0	0
[65-70)	0	0	0	0	0	0
[70-75)	0	0	0	0	0	0
[75-80)	0	1	2	0	3	1
[80-85)	0	0	1	0	1	0
[85-90)	0	0	0	0	0	0
[90-95)	0	0	0	0	0	0
[95-100)	0	0	0	0	0	0
[100-105]	0	0	0	0	0	0
Total	137	151	104	77	469	102

For each gender, the distribution of the number of deaths has been estimated by applying the proportion for each age group, derived from the column headed “Sample 1837-42” shown in Table 2.15.5, to the estimated total number of deaths for 1841 (*see* Table 2.14.7). The estimated distribution is given in Table 2.15.5.

The estimated distributions of population size, and the estimated distributions of the number of deaths, have been tabulated using the age groupings for a conventional abridged life table analysis to produce the two data files for this section. To conform to the data specifications of Section 2.3, the estimated number of deaths for each age group shown in Table 2.15.5 is included in triplicate in these data files. The data is internally consistent, in the sense that at each age the number of deaths is not greater than the population size. To be consistent with previous sections, the average value of each of the five-year age groupings is used as the single-year value within that group as an approximation for complete life table analysis, for both population and mortality distributions. The two data files and the four output files produced by the thesis computer program are included on the CD-rom using the file naming conventions of Section 2.3.

Table 2.15.6 (*cf* Table 2.6.3 and Table 2.6.4) shows estimates of current expected life with standard error for each gender, calculated from abridged life tables using the q-method. *Z*-scores are used to compare the estimates of current expected life between genders, and the estimate of current expected life for males is not statistically significantly different from the estimate for females. The expected values of the length of life at birth for England for 1841 are repeated in Table 2.15.6. The estimate for males for England is not statistically significantly different from the thesis estimate for South Australia, while the thesis estimate for females for South Australia is statistically significantly higher than the corresponding English estimate. For z-score calculations, it is assumed that the error of the English estimate is equal to the error of the corresponding thesis estimate.

Table 2.15.6: Estimates of current expected life for 1841

Census Year	Male	Female	Z-score
1841	44.27 (2.59)	48.17 (1.88)	-1.2
ELT 1841	40.0		1.1
		42.0	2.2

Table 2.15.7 (*cf* Table 2.6.5) shows estimates of current expected life with standard error for each combination of life table type and method of analysis, for each census year and gender. Table 2.15.7 also shows estimates designated as "Doering", discussed in detail in Section 2.11 and Section 2.13. These estimates have been calculated from a tabulation of the population and mortality data using the reported age groupings of [0-7), [7-14), [14-21), [21-35) and [35-50) of the census population counts, combined with standard five-year groupings of [50-55), [55-60),…[100-105]. For each gender, the estimates from a conventional abridged current life table and from a "Doering" current life table are not statistically significantly different (z-scores, not shown).

Table 2.15.7: Methodological comparisons for 1841

Life Table Type	Male		Female	
	q-method	u-method	q-method	u-method
complete	43.78 (1.56)	43.50 (1.49)	47.47 (1.43)	47.57 (1.41)
abridged	44.27 (2.59)	43.87 (2.26)	48.17 (1.88)	48.24 (1.90)
Doering	42.00 (2.61)		46.22 (2.17)	

Table 2.15.8 (*cf* Table 2.10.5) shows the results extracted from the various output files to illustrate the potential effect of undercount of reported census population counts and number of deaths on the estimates of current expected life. The extent of estimation that has been undertaken for both population size and number of deaths may limit the interpretational value of this table. It has been included for completeness with other sections.

Table 2.15.8: Robustness comparisons for 1841

	Standard	Population+5%	Deaths+10%	Population+10%
Male	44.27 (2.59)	45.60 (2.57)	41.60 (2.40)	46.82 (2.62)
Female	48.17 (1.88)	49.29 (1.86)	46.11 (1.85)	50.27 (1.78)

2.16 Summary and discussion

The estimates of current expected life and corresponding standard error calculated in Section 2.6 to Section 2.15 are shown in Table 2.16.1 for each census year and gender. The estimates shown are those obtained from an abridged life table using the q-method prior to and including 1901, and from a complete life table using the q-method from 1911. The total population size and average total number of deaths are also shown.

Table 2.16.1: Current expected life with standard error for the period 1841-1996

	Male				Female			
Census Year	**Population**	**Deaths**	**CEL**	**SE(CEL)**	**Population**	**Deaths**	**CEL**	**SE(CEL)**
1841	8195	140	44.27	2.59	6345	102	48.17	1.88
1844	9525	172	41.95	1.56	7608	128	43.49	1.60
1846	12670	236	41.57	1.55	9650	163	46.67	2.08
1851	35309	544	45.73	1.24	27730	439	47.62	1.31
1855	43716	713	47.68	1.27	41470	620	50.00	1.27
1861	64640	1111	48.55	1.06	61678	951	51.88	1.20
1866	85625	1434	47.39	0.86	77975	1178	50.34	0.88
1871	95288	1438	48.69	0.67	90164	1165	52.19	0.77
1876	110410	1972	43.57	0.58	102734	1659	46.09	0.64
1881	149530	2283	47.62	0.50	130335	1819	50.76	0.62
1891	166801	2156	52.52	0.47	153630	1787	54.98	0.52
1901	184424	2169	54.56	0.47	178182	1821	57.72	0.50
1911	207358	2269	58.81	0.41	201200	1880	62.41	0.42
1921	248267	2663	60.01	0.38	246893	2221	63.53	0.39
1933	290429	2701	65.34	0.34	289546	2386	68.04	0.35
1947	320031	3478	67.17	0.27	326042	2995	71.28	0.26
1954	403903	3958	67.72	0.25	393191	3267	73.07	0.24
1961	490225	4412	68.62	0.22	479115	3537	74.91	0.21
1966	548530	5030	68.57	0.21	543344	4034	75.22	0.20
1971	586451	5454	68.67	0.20	587656	4408	75.58	0.20
1976	619759	5402	70.06	0.19	624595	4488	76.93	0.18
1981	635695	5473	71.49	0.19	649467	4438	78.92	0.18
1986	666159	5626	72.98	0.18	679985	4824	79.48	0.18
1991	690805	5856	74.08	0.17	709802	5156	80.30	0.17
1996	698799	5989	75.33	0.17	722673	5504	81.34	0.16

Figure 2.16.1 displays graphically the change over time of the estimates of current expected life for each gender. Several events that may aid in the interpretation of this graph are indicated as

- 1933 and 1939, with larger time periods between estimates related to the occurrence of the Great Depression and World War II.
- 1901, the census year marking the separation of estimates based on abridged life tables to estimates based on complete life tables. In Section 2.6 to Section 2.9, population and mortality data is available both at the single-year level, for complete life table analysis, and, after appropriate amalgamation of this single-year data, as grouped data for abridged life table analysis. The corresponding estimates of current expected life using the q-method are essentially equal. It is assumed that this "interchangeability" between estimates is a reasonable assumption for the period 1841-1901, where generally the available data is tabulated in a form primarily suitable for abridged life table analyses rather than complete life table analyses.
- 1876 and 1881, where the generally increasing pattern in the estimated values for both genders is disrupted. This is an illustration of the estimator property of current expected life to summarise the prevailing mortality of the calendar year at which the estimate is calculated. In the 1870s the societal and health conditions of Adelaide, called by some commentators "the city of stenches", had degenerated with frequent epidemics of infectious diseases such as measles, malaria and typhoid fever. Government action was urgently needed. The first Public Health Act was passed in 1873, and systems of deep drainage and sewerage works were introduced into Adelaide and the surrounding townships by 1880 [56], [57].
- 1861, the first census year at which the estimate of current expected life for males is statistically significantly different from the estimate for females (determined by z-score).

The estimates of current expected life for 1841, 1844 and 1846 obviously have to be interpreted cautiously, principally because of the inferred mortality distributions for these years (*see* Table 2.14.7). If the expected lifetimes from the ELT1841 are considered as appropriate reference values, then the thesis estimates for these three years are too high, particularly for females. This suggests an underestimation of the number of deaths. However, in conjunction with the estimated standard errors, the thesis estimates are not entirely unbelievable. Moreover, for each gender, they are essentially no larger than the estimate for 1851, the first year for which there are observed population and mortality distributions.

Figure 2.16.2 displays the relationship between the estimated standard error and the total population size on which the estimate was based, combined over gender but excluding the results for 1841, 1844 and 1845. This graph indicates an approximate linear relationship between standard error and the inverse of the square root of total population size, reflecting a common statistical property of variance being inversely proportional to sample size. A simple linear regression line was fitted to the points and is also shown in Figure 2.16.2, with parameter estimates of the line given in Table 2.16.2.

Table 2.16.2: Parameters of regression lines for standard error on population size

Equation : SE = a + b / $\sqrt{\text{Population size}}$			
	a	**b**	
Both genders (Figure 2.16.2)	-.179	281.669	
Male only (not displayed)	-.191	285.878	(Test of equality of
Female only (not displayed)	-.168	278.099	regression lines; p=.87)

The estimated standard errors that have the largest deviations from the fitted regression line are indicated in Figure 2.16.2 by the census year to which they correspond, specifically 1851, 1855 and 1861. The lack of conformity to the empirical prediction rule provided by the regression line is most probably an effect of the extensive use of the imputation procedure (*see* Section 1.9) within the bootstrap estimation process of current expected life, for these years in particular.

Figure 2.16.1: Current expected life for the period 1841-1996

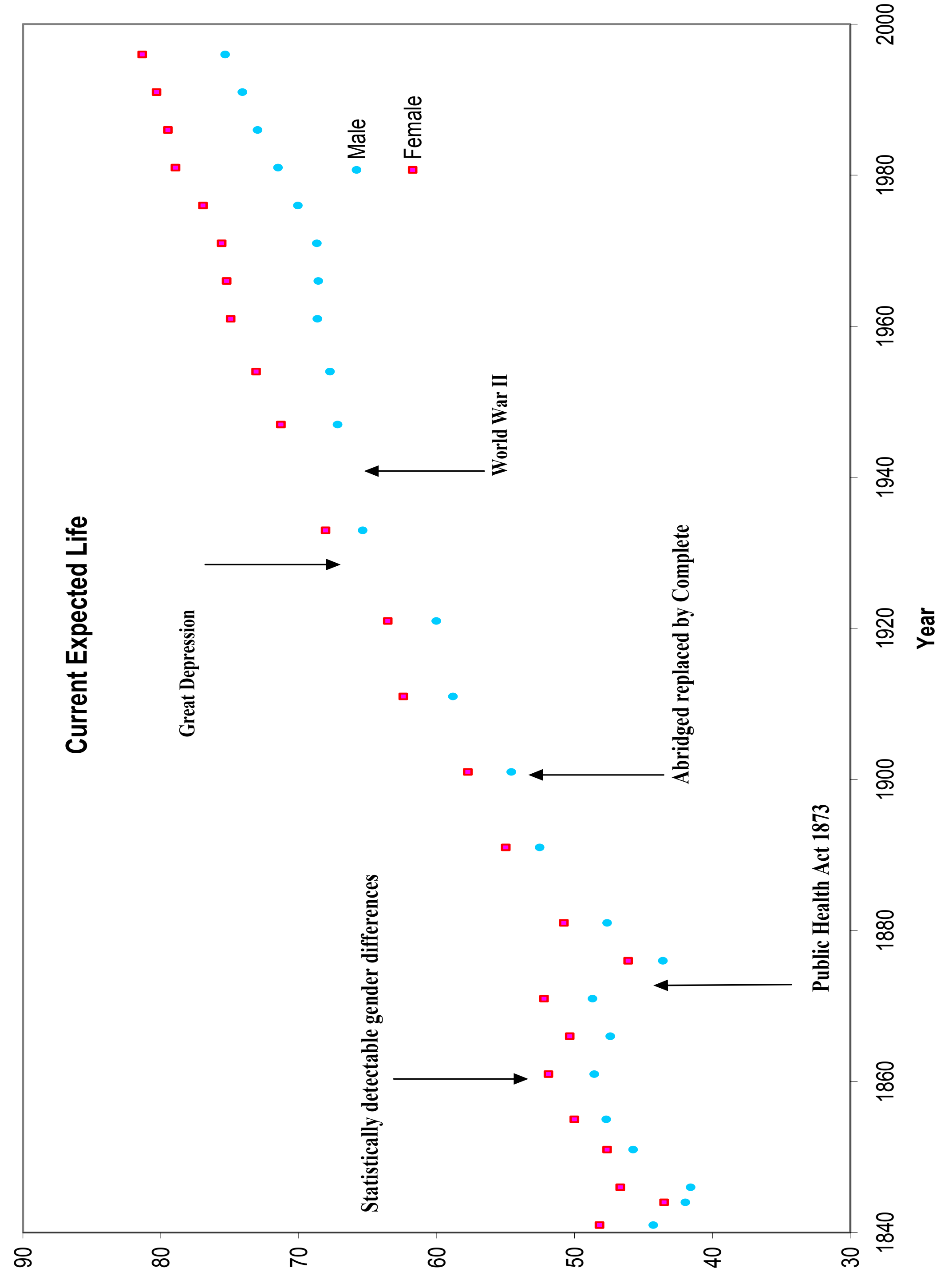
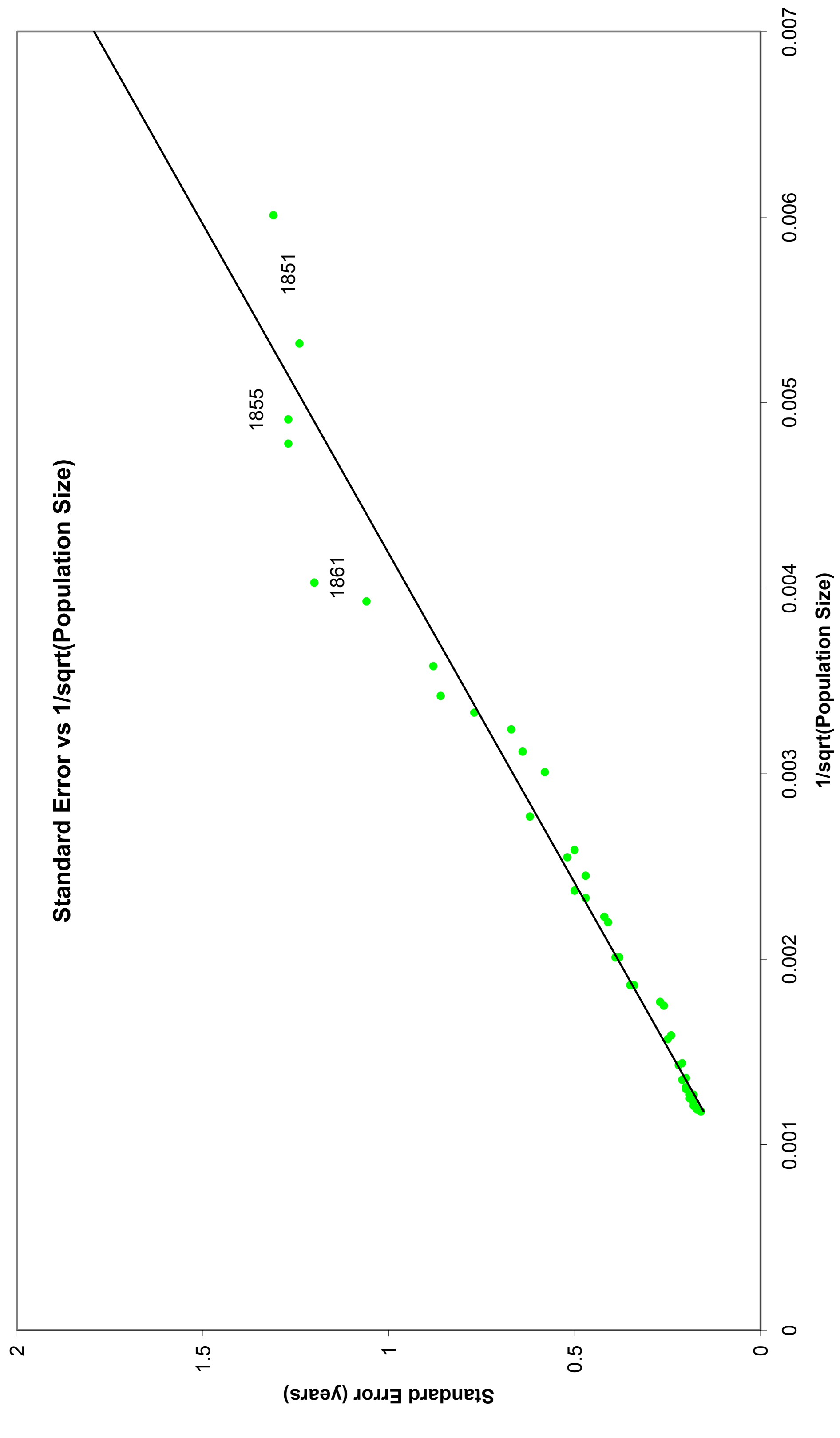


Figure 2.16.2: The relationship between standard error and total population size



CHAPTER 3: GENERATION EXPECTED LIFE FOR SOUTH AUSTRALIA

3.1 Introduction

The concept of current expected life for a particular calendar year is established in Chapter 1 through a distribution function that is synthesised from a sequence of lifetime distribution functions, where each function of the sequence is specific to a calendar year previous to the designated year of interest. For example, it is shown in Section 1.7 that one form of current expected life, determined from a complete life table analysis using the q-method, is derived from the synthesised distribution function G_C in which

$$G_C(k+1) = 1 - [1 - \frac{F_0(1) - F_0(0)}{1 - F_0(0)}]\,[1 - \frac{F_1(2) - F_1(1)}{1 - F_1(1)}]....[1 - \frac{F_k(k+1) - F_k(k)}{1 - F_k(k)}] \text{ for } k = 0, (1), 104$$

and where F_0 is the lifetime distribution function applicable to individuals born in the calendar year of interest; F_1 is the lifetime distribution function applicable to individuals born in the first year prior to the calendar year of interest; F_2 is the lifetime distribution function applicable to individuals born in the year two years prior to the calendar year of interest, and so on, moving backwards in time from the primary year. It is further suggested in Section 1.7 that

- $G_C \neq F_0$, unless $F_0 = F_1 = ... = F_{105}$
- the expected lifetime calculated from G_C is less than or equal to the expected lifetime calculated from F_0, under reasonable assumptions

Similar comments apply to the other synthesised distribution function G_A that is also defined in Chapter 1.

For these reasons, current expected life is often considered a “statistical fiction” (Greenwood [58]); a value that summarises the idealised mortality of a hypothetical population rather than the actual mortality of an identifiable group of real individuals. Consequently, care needs to be taken in the interpretation of any estimate of current expected life if it is used as a prediction of future lifetime.

In contrast, and for some circumstances, data sets similar to those analysed individually in Chapter 2 can be used collectively to estimate the lifetime distribution function F_0 pertinent to the actual population of a specific calendar year. The approximation of F_0 obtained in this way is called the generation lifetime distribution function, and generation expected life, an estimate of the expected value of the distribution F_0, is obtained from this function. The principles are outlined briefly at this stage, with formal mathematical development given in Section 3.2 and Section 3.3.

Conceptually, pertinent data sets of census population counts and associated number of deaths are arranged in ascending chronological order by calendar year of census. These censuses preferably span a time period of at least one hundred years. The earliest census year of the ordered group is taken as the year of birth of the individuals who comprise the "generation", and thus determines the particular function F_0 that is estimated. The population count of individuals aged less than one year from this first census, and the resulting mortality rate, can refer only to individuals of the "generation" in their first year of life. Clearly, this earliest mortality rate is related in some way to a combination of functional values of F_0. The details of this relationship are discussed in Section 3.1. Suppose now that the second census occurs five years after the first, by which time the "generation" will be five years old. In the absence of migration, the population count in the five-year age group from this second census again can only be of individuals from the "generation". Hence the mortality rate derived for the five-year age group from the second census is also related to functional values of F_0. Similarly, if the third census is taken after a further five years, then the "generation" will be ten years old, and the mortality rate derived for this age group from the third census is related to yet another combination of functional values of F_0. Continuing in this manner, it can be seen that the aging of the "generation", relative to the occurrence of censuses subsequent to the first, determines a selection of mortality rates that can be combined to approximate F_0 over the "generation" lifespan, being the time between the first and the last census. Of course, both immigration and emigration would happen to the "generation" over this extended period of time for any real population of individuals. However, assuming that this interchange takes place with external populations that have equivalent lifetime distribution functions to the "generation", no effect of any consequence is anticipated on the interpretation of this estimation of F_0.

Generation expected life can only be estimated retrospectively because data is needed over the entire lifespan of the "generation" to establish the generation lifetime distribution function. This requirement places the year of birth of any "generation" approximately one hundred years earlier than the year of collection of the last necessary and relevant data. For this reason, and in contrast to current expected life, it is therefore not possible to use generation expected life as a predictor of future lifetime. However, for biological or demographic assessment, generation expected life is the only practical way to satisfactorily quantify the realised mortality of a large, unconstrained and freely-living human population.

There are two major studies reporting generation expected life for Australia, and while both studies incorporate some South Australian data into Australian estimates, generation expected life is not estimated specifically for South Australia in either study. Lancaster [59] gives generation expected life for each gender for the years 1851,(5),1891. His calculations are based on an unequal mixture of data from all Australian colonies and States, with data from the earliest years being predominantly from New South Wales, as reported by Pell [45], and which Lancaster assumes "as good estimates for Australia". South Australian data for the years after 1870 is included in his study, and the difficulties with South Australian data discussed in detail in Section 2.11 are resolved by Lancaster for his purposes by using broad distributional approximations. Young [60] gives current and generation expected life for each gender for the years 1850,(10),1960 in a doctoral thesis on Australian population growth and mortality. In marked contrast to both Lancaster and this present thesis, Young constructs the population age distributions at certain calendar years from the birth and death records of preceding years, maintaining that the population age distributions derived in this way are more reliable than the corresponding census population counts. Young also states that it "would be virtually impossible to construct cohort [*i.e.* generation] life tables according to the mortality rates for each State, as has been done for Australia" but warns that "before 1880, the concept of "average" Australian experience could cover a wide range of probabilities of survival". Both Lancaster and Young, particularly, provide references to the extensive literature on generation expected life.

3.2 Rationale for a generation lifetime distribution function

The detailed argument for the development of a generation lifetime distribution function is presented for ease of explanation by reference to a particular situation. The following derivation uses the notation of Section 1.5, which is extended in this section to indicate functions pertinent to a specific calendar year.

Firstly, consider the census year 1901. Let ${}_1D_0^{1901}$ denote the expected number of deaths between birth and one year occurring in 1901, and let ${}_1P_0^{1901}$ denote the expected number alive aged less than one year at the time of the census.

Applying the results of Section 1.5,

$$q_0^{1901} = {}_1q_0^{1901} = {}_1D_0^{1901} / ({}_1P_0^{1901} + \tfrac{1}{2}\,{}_1D_0^{1901}) \approx \frac{F_0(1) - F_0(0)}{1 - F_0(0)} = \frac{F^{1901}(1) - F^{1901}(0)}{1 - F^{1901}(0)}$$

in which the notation F_0, the lifetime distribution function of individuals born in the calendar year 1901 on which the argument is currently focused, is substituted by the equivalent, but more explicit, notation F^{1901}.

Consider now the following year 1902, and let ${}_1D_1^{1902}$ denote the expected number of deaths occurring between one and two years of age in this calendar year. Although 1902 is a non-census year, let ${}_1P_1^{1902}$ denote the expected number alive aged between one and two years at a hypothetical census taken at any time within this year. Again with reference to the results of Section 1.5

$$q_1^{1902} = {}_1q_1^{1902} = {}_1D_1^{1902} / ({}_1P_1^{1902} + \tfrac{1}{2}\,{}_1D_1^{1902}) \approx \frac{F_1(2) - F_1(1)}{1 - F_1(1)} = \frac{F^{1901}(2) - F^{1901}(1)}{1 - F^{1901}(1)}$$

The final part of this expression follows since in this instance, within the context of the calendar year 1902 and by the notation defined in Section 1.5, F_1 refers to the lifetime distribution function of individuals born in the year immediately prior to 1902.
Hence $F_1 = F^{1901}$.

Similar treatment of the successive years 1903, 1904 *etc*, produces the expressions $q_2^{1903} \approx \frac{F^{1901}(3) - F^{1901}(2)}{1 - F^{1901}(2)}$, $q_3^{1904} \approx \frac{F^{1901}(4) - F^{1901}(3)}{1 - F^{1901}(3)}$ and so on, which generalizes as

$$q_k^{1901+k} \approx \frac{F^{1901}(k+1) - F^{1901}(k)}{1 - F^{1901}(k)} \text{ for } k = 0, 1, 2, 3\ldots.$$

The generation lifetime distribution function $\mathcal{G}^{1901}$ is then defined in which

$$\mathcal{G}^{1901}(k+1) = 1 - (1-q_0^{1901})\,(1-q_1^{1902})\ \ldots\ (1-q_k^{1901+k}) = 1 - \prod_{i=0}^{k}(1-q_i^{1901+i}) \quad \text{for } k = 0, 1, 2, 3\ldots.$$

Substituting the approximations above and by identification with Section 1.3.1 gives $\mathcal{G}^{1901}(k+1) \approx F^{1901}(k+1)$ for $k = 0, 1, 2, 3\ldots.$

Since this reasoning readily applies to any calendar year of interest, the generation lifetime distribution function for calendar year Y, $\mathcal{G}^{Y}$, is thus defined in which

$$\mathcal{G}^{Y}(k+1) = 1 - \prod_{i=0}^{k}(1-q_i^{Y+i}) \approx F^{Y}(k+1) \text{ for } k = 0, 1, 2, 3$$

The generation lifetime distribution function $\mathcal{G}^{Y}$ is analogous to the current lifetime distribution function G_C that is defined in Section 1.7, in the sense that each function is based on the same arithmetical combination of the same expected values that are pertinent to a time unit of one year. Similarly, it is possible to produce generation counterparts to the other current lifetime distribution functions G_A, H_C and H_A, also defined in Section 1.7, but these variations are not considered in this thesis.

3.3 Estimation of generation expected life

The procedure that is used to estimate values for the quantities q_i^y, which define the generation distribution function $\mathcal{G}^{Y}$ for calendar year Y, is illustrated using the specific example of $\mathcal{G}^{1901}$.

The quantities q_i^y, i = 0,(1),10 and y = 1901,(1),1911, are displayed in the following schematic as a rectangular array, with columns corresponding to calendar year. The entries shown with a shaded background are the first eleven components of $\mathcal{G}^{1901}$ for which numerical estimates are needed. The first and last columns, for the census years 1901 and 1911, are shown in bold font to indicate that all quantities in these two columns can be estimated directly by the method used for current life table analysis and given in Section 1.9. For the intervening non-census years of 1902 to 1910, direct estimation of any of the quantities in these columns, and of the shaded elements in particular, is not possible from routinely collected data. Assuming a regularity of change over the years between censuses, linear interpolation (*see* Section 1.4.3) between the estimated values of $\mathbf{q}_{\mathbf{i}}^{\mathbf{1901}}$ and $\mathbf{q}_{\mathbf{i}}^{\mathbf{1911}}$ is used to impute values for q_i^y, y = 1902,(1),1910 , for i = 2,(1),9.

$\mathbf{q}_{\mathbf{0}}^{\mathbf{1901}}$	q_0^{1902}	q_0^{1903}	q_0^{1904}	q_0^{1905}	q_0^{1906}	q_0^{1907}	q_0^{1908}	q_0^{1909}	q_0^{1910}	$\mathbf{q}_{\mathbf{0}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{1}}^{\mathbf{1901}}$	q_1^{1902}	q_1^{1903}	q_1^{1904}	q_1^{1905}	q_1^{1906}	q_1^{1907}	q_1^{1908}	q_1^{1909}	q_1^{1910}	$\mathbf{q}_{\mathbf{1}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{2}}^{\mathbf{1901}}$	q_2^{1902}	q_2^{1903}	q_2^{1904}	q_2^{1905}	q_2^{1906}	q_2^{1907}	q_2^{1908}	q_2^{1909}	q_2^{1910}	$\mathbf{q}_{\mathbf{2}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{3}}^{\mathbf{1901}}$	q_3^{1902}	q_3^{1903}	q_3^{1904}	q_3^{1905}	q_3^{1906}	q_3^{1907}	q_3^{1908}	q_3^{1909}	q_3^{1910}	$\mathbf{q}_{\mathbf{3}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{4}}^{\mathbf{1901}}$	q_4^{1902}	q_4^{1903}	q_4^{1904}	q_4^{1905}	q_4^{1906}	q_4^{1907}	q_4^{1908}	q_4^{1909}	q_4^{1910}	$\mathbf{q}_{\mathbf{4}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{5}}^{\mathbf{1901}}$	q_5^{1902}	q_5^{1903}	q_5^{1904}	q_5^{1905}	q_5^{1906}	q_5^{1907}	q_5^{1908}	q_5^{1909}	q_5^{1910}	$\mathbf{q}_{\mathbf{5}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{6}}^{\mathbf{1901}}$	q_6^{1902}	q_6^{1903}	q_6^{1904}	q_6^{1905}	q_6^{1906}	q_6^{1907}	q_6^{1908}	q_6^{1909}	q_6^{1910}	$\mathbf{q}_{\mathbf{6}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{7}}^{\mathbf{1901}}$	q_7^{1902}	q_7^{1903}	q_7^{1904}	q_7^{1905}	q_7^{1906}	q_7^{1907}	q_7^{1908}	q_7^{1909}	q_7^{1910}	$\mathbf{q}_{\mathbf{7}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{8}}^{\mathbf{1901}}$	q_8^{1902}	q_8^{1903}	q_8^{1904}	q_8^{1905}	q_8^{1906}	q_8^{1907}	q_8^{1908}	q_8^{1909}	q_8^{1910}	$\mathbf{q}_{\mathbf{8}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{9}}^{\mathbf{1901}}$	q_9^{1902}	q_9^{1903}	q_9^{1904}	q_9^{1905}	q_9^{1906}	q_9^{1907}	q_9^{1908}	q_9^{1909}	q_9^{1910}	$\mathbf{q}_{\mathbf{9}}^{\mathbf{1911}}$
$\mathbf{q}_{\mathbf{10}}^{\mathbf{1901}}$	q_{10}^{1902}	q_{10}^{1903}	q_{10}^{1904}	q_{10}^{1905}	q_{10}^{1906}	q_{10}^{1907}	q_{10}^{1908}	q_{10}^{1909}	q_{10}^{1910}	$\mathbf{q}_{\mathbf{10}}^{\mathbf{1911}}$

The array shown above is a small part of a conceptually larger array, which expands to the right and below by the addition of further quantities q_i^y with i > 10 and y > 1911. The quantities of interest needed for $\mathcal{G}^{1901}$ continue to be the diagonal entries of this enlarged array. The estimation procedure described above extends naturally; taking as the estimate of q_i^y either a value calculated from the observations for a census year or an appropriately interpolated value for a non-census year.

This estimation procedure readily generalizes to any initial calendar year Y, census or non-census, and the estimated values for $q_i^y = q_i^{Y+i}$ are used in the expression for $\mathcal{G}^Y$ given in Section 3.2 to produce an estimate of this function, denoted by $\hat{\mathcal{G}}^Y$.

The estimate of generation expected life is calculated by using $\hat{\mathcal{G}}^Y$ and the numerical methods described in Section 1.3.3 and Section 1.4.2.

3.4 The thesis FORTRAN computer program: << generation.f >>

The procedures indicated in Section 3.3 for calculating an estimate of generation expected life for calendar year Y have been programmed as a thesis FORTRAN computer program with a source code file << generation.f >> (*included* CD-rom) and an executable file << generation.exe >> (*included* CD-rom). The program is executed in a MS-DOS window by the command line statement "generation < generationin.txt > generationout.txt".

The standard text file << generationin.txt >> is a listing of the input specifications for this program and consists of one or more lines of data. Each line is arranged as

$$\text{"}Y, yyyy_1, yyyy_2, \ldots, yyyy_{ny}\text{"}$$

where $yyyy_1, yyyy_2, \ldots, yyyy_{ny}$ specifies census years with $yyyy_1 < yyyy_2 < \ldots < yyyy_{ny}$. It is implied that $yyyy_1 \leq Y$, and that the specified census years span the lifetime of the generation. It is assumed that all data files (*see* Section 2.3) of the form << M $yyyy_i$ .txt >> and << F $yyyy_i$ .txt >> for i = 1, ..,ny are accessible at execution.

The standard text file << generationout.txt >> contains the bootstrap estimates (*see* Section 1.11.3) of generation expected life for males and females for each input value of $yyyy_1$. The bootstrap standard error of the estimate of generation expected life is also included, as are the 5^{th}, 50^{th} and 95^{th} percentiles of the bootstrap distribution of generation expected life.

The bootstrap procedure adapted for generation expected life has been implemented only to produce random variation in the observed deaths for the nominated census years, *i.e.* in the notation of Section 1.11.3, using the specification δ = (1,0,1). As a result variation is induced into the values linearly interpolated between censuses. Finally, all estimates of q_i^y result from y constrained to be the minimum of ($yyyy_1$ + i) and $yyyy_{ny}$.

3.5 Some estimates of generation expected life for South Australia

For each gender, Table 3.5.1 shows bootstrap estimates of generation expected life and standard error for the years 1851, 1881 and 1901 calculated by the thesis computer program described in Section 3.4. This table also includes the corresponding estimate of current expected life reported in Chapter 2, and gives the ratio of the magnitudes of the two types of estimate.

The comparisons of current expected life and generation expected life included in Table 3.5.1 illustrate the potential for error when current expected life is used as a predictor of future lifetime. For 1851, the estimates for each type are essentially the same for both genders. In this case the general assumption, that future mortality will be no more than current mortality, does not hold uniformly in the years after 1851. The irregular mortality conditions in the years following 1851 are incorporated into the estimates used to calculate generation expected life, but are not available, by definition, as estimates for the calculation of current expected life (*see* also Figure 2.16.1). For 1881 and 1901, the estimate of generation expected life indicates that the corresponding estimate of current expected life under-predicts the experienced lifetime of those generations by an amount in the range 7% to 12%.

Table 3.5.1: Estimates of generation expected life for 1851, 1881 and 1901

	Male		
	1851	**1881**	**1901**
Census years included in the estimation of GEL	1851,1855,1861,1866 1871,1876,1881,1891 1901,1911,1921,1933 1947,1954,1961	1881,1891,1901,1911 1921,1933,1947,1954 1961,1966,1971,1976 1981,1986,1991	1901,1911,1921,1933 1947,1954,1961,1966 1971,1976,1981,1986 1991,1996
Generation Expected Life (SE)	45.08 (.71)	51.70 (.40)	58.53 (.39)
Ratio Generation : Current	0.99	1.09	1.07
5th Percentile	43.90	51.05	57.90
50th Percentile	45.06	51.69	58.54
95th Percentile	46.28	52.36	59.17
Current Expected Life (SE)	45.73 (1.24)	47.62 (.50)	54.56 (.47)

	Female		
	1851	**1881**	**1901**
Census years included in the estimation of GEL	1851,1855,1861,1866 1871,1876,1881,1891 1901,1911,1921,1933 1947,1954,1961	1881,1891,1901,1911 1921,1933,1947,1954 1961,1966,1971,1976 1981,1986,1991	1901,1911,1921,1933 1947,1954,1961,1966 1971,1976,1981,1986 1991,1996
Generation Expected Life (SE)	48.61 (.74)	55.49 (.43)	64.58 (.41)
Ratio Generation : Current	1.02	1.09	1.12
5th Percentile	47.38	54.78	63.91
50th Percentile	48.62	55.49	64.58
95th Percentile	49.85	56.19	65.26
Current Expected Life (SE)	47.62 (1.31)	50.76 (.62)	57.72 (.50)

3.6 The effect of the events of 1914-19 on male generation expected life 1881-1900

In Section 3.3, linear interpolation is presented as a method to estimate a value for any quantity q_i^y for a non-census year y. The suitability of this method relies on the assumption of regularity in change of the notional q_i^y values between any two successive census years that are used as a basis for the interpolation process. In some periods of time there may be extraordinary events that seriously invalidate this assumption. Estimates of generation expected life of South Australian males for 1881-1900 are used as an illustration of this assumption.

The generation lifetime distribution function for each of the calendar years in the period 1881-1900 have a number of component q-values that coincide with the calendar years 1915-19, and these coincident values are indicated in Table 3.6.1.

For each generation lifetime distribution function, the first component of the relevant set of components corresponds with, at least, the legal age of eighteen at which an individual born in the generation could serve in the armed forces during the years of World War I. For the purposes of the calculations presented in this section, each component is taken to represent the composite mortality at each age of those who did not enlist and remained in South Australia, and those who enlisted and served active military duty. The last component of each set coincides with the influenza pandemic of 1918-19 [61]. It is reported in the *Annual Return of the RBDM for1920* that in South Australia in 1919 " 532 deaths (303 males and 229 females) were directly caused by the epidemic, and some hundreds more in which some other disease was registered as the cause, but which were attributable to the effects of influenza". In contrast, the total number of deaths attributed to influenza in 1918 and 1920 in South Australia was 47 and 24 respectively.

It was because of the occurrence of these two major disruptive events that the Australian Life Tables for 1921 were based on the numbers of deaths that occurred in the years 1920-22; a change from the methodology previously applied that would have used the numbers of deaths over the ten-year period between the censuses of 1911 and 1921. The *Statistician's Report on the Census of 1921* says that to do otherwise would result "in a combination of rates not likely to be experienced in the near future."

Table 3.6.1 Components of generation lifetime distribution functions 1881-1900

Generation	Calendar years and q-value component of generation distribution function				
	1915	1916	1917	1918	1919
1881	q_{34}	q_{35}	q_{36}	q_{37}	q_{38}
1882	q_{33}	q_{34}	q_{35}	q_{36}	q_{37}
1883	q_{32}	q_{33}	q_{34}	q_{35}	q_{36}
1884	q_{31}	q_{32}	q_{33}	q_{34}	q_{35}
1885	q_{30}	q_{31}	q_{32}	q_{33}	q_{34}
1886	q_{29}	q_{30}	q_{31}	q_{32}	q_{33}
1887	q_{28}	q_{29}	q_{30}	q_{31}	q_{32}
1888	q_{27}	q_{28}	q_{29}	q_{30}	q_{31}
1889	q_{26}	q_{27}	q_{28}	q_{29}	q_{30}
1890	q_{25}	q_{26}	q_{27}	q_{28}	q_{29}
1891	q_{24}	q_{25}	q_{26}	q_{27}	q_{28}
1892	q_{23}	q_{24}	q_{25}	q_{26}	q_{27}
1893	q_{22}	q_{23}	q_{24}	q_{25}	q_{26}
1894	q_{21}	q_{22}	q_{23}	q_{24}	q_{25}
1895	q_{20}	q_{21}	q_{22}	q_{23}	q_{24}
1896	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
1897	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}
1898		q_{18}	q_{19}	q_{20}	q_{21}
1899			q_{18}	q_{19}	q_{20}
1900				q_{18}	q_{19}

For each generation, clearly an estimate of any q-value obtained by linear interpolation between $\mathbf{q_i^{1911}}$ and $\mathbf{q_i^{1921}}$, both calculated from data associated with the peacetime census years of 1911 and 1921, is a conservative measure of the true mortality within the period 1915-19. Conversely, it is assumed for this thesis that the other q-values defining the generation lifetime distribution function and not included in the specified set can be adequately estimated by linear interpolation between relevant census years. For example, individuals of the generation of males born in 1900 would be 39 years old by 1939 and the beginning of World War II. Of this particular generation, it is considered that relatively few would have enlisted and thus be exposed to increased mortality due to active war-time conditions. Hence q_{38}^{1939} is assumed to be reasonably estimated by linear interpolation between $\mathbf{q_{38}^{1933}}$ and $\mathbf{q_{38}^{1947}}$.

Table 3.6.2 shows the average number of male deaths registered in South Australia for ages 18 to 41 for 1915-19, calculated from figures published in ABS *Demography Australia* 1915-20. These averages are based explicitly on civilian registrations of deaths, and do not include war-related fatalities. The effect of the Influenza Pandemic can be seen by the increase in the age-specific average number of deaths for 1919.

Table 3.6.2: Average number of civilian deaths per year for South Australian males 1915-19

Ages	1915-18	1919
18-19	11	18
20-25	13	16
26-30	15	29
31-35	17	33
36-41	19	31

A tabulation by age and year of death of the number of deaths as a direct consequence of military service is not readily available. Neither the official history of the Great War by Bean [62], nor an extensive compilation by the Records Section of the 1st Australian Imperial Force (1/AIF) [63] provides the necessary statistical detail required for this thesis. The focus of these publications is on the larger and more immediate issues of campaigns, military operations and troop disposition. It may never be possible to accurately determine this level of detail owing to, as Butler says in the history of the Australian Army medical services [64], "the accidental destruction of part of the Australian records by the British Ministry of Pensions and Office of Works [and that] many results have to be arrived at by computation from sample counts". Locally, a detailed history of the 10th Battalion 1/AIF [65], raised in South Australia in 1914, lists the name and date of death of each of the 1010 fatalities experienced by this battalion over the course of the war, but does not include the age at death.

An attempt has been made for this thesis to compile from primary sources the necessary statistical information needed to calculate more realistic estimates of generation expected life than those calculated by using the interpolation procedure. However, in attempting this task it is acknowledged that, as Gammage [66] says, "Great War statistics are notoriously variable" and that "These [sources] have various and conflicting statistics".

The Australian War Memorial, Canberra, maintains a publicly accessible internet site [http://www.awm.gov.au/database/roh.asp, August 2002] that includes a comprehensive and extensive Roll of Honour for the known Australian war dead from all conflicts in which Australia has been involved. A key-word search of "+SA", used in combination with the specification "World War 1914-1918" in the "Conflict" field, produced a listing of 5756 service records. In this search, matching could have occurred through an entry of "SA" in one or more of the record fields "Place of death", "Cemetery", "Place of enlistment" or "Native place". The Roll of Honour is organized to display individual service records in groups of ten, and the 5756 identified records were downloaded to a computer text file suitable for character manipulation and data extraction by a repetitive "request next ten records / select / copy / paste" procedure. Unfortunately for the present purposes, an individual service record does not have a specific field for age at death. This item of information is often but not always included in a "Notes" field that was sometimes added to the Roll of Honour to provide additional biographical detail of an individual. A computer program has been written which isolates each service record from the downloaded text file, searches the data fields, and determines where possible the age and year of death of the individual. The results of this extraction process are summarized in Table 3.6.3. It was possible to obtain both age and year of death for 3446 or approximately 60% of the listed 5756 individuals.

Table 3.6.3: Number of war deaths of South Australian males from the Australian War Memorial Roll of Honour

Age	Year of Death						
	Unstated	1915	1916	1917	1918	1919	Total
Unstated	58	341	603	726	510	67	2305
16	0	0	2	1	0	0	3
17	0	0	4	3	0	0	7
18	0	6	25	17	4	0	52
19	0	26	59	61	39	1	186
20	1	37	69	88	64	2	261
21	1	27	65	96	51	2	242
22	0	24	78	128	81	2	313
23	0	19	78	116	92	2	307
24	2	25	67	106	67	4	271
25	0	22	52	103	64	2	243
26	0	17	56	85	49	0	207
27	0	22	28	78	56	5	189
28	0	15	32	58	50	6	161
29	0	12	37	47	48	0	144
30	0	12	33	46	40	4	135
31	0	8	24	42	35	0	109
32	0	9	21	40	42	2	114
33	0	10	19	37	27	2	95
34	0	7	15	16	29	1	68
35	0	7	10	26	15	2	60
36	0	4	14	17	16	2	53
37	0	5	9	10	13	1	38
38	0	2	11	12	19	0	44
39	0	3	9	12	7	0	31
40	0	2	12	7	7	0	28
41	0	1	8	8	4	0	21
42	1	2	2	6	4	1	16
43	0	0	4	3	4	1	12
44	0	1	2	4	1	0	8
45	0	3	6	4	2	0	15
46	0	1	2	3	3	0	9
47	0	0	1	0	1	0	2
48	0	0	0	2	0	0	2
49	0	0	0	1	1	0	2
50+	0	0	2	0	1	0	3
Total	63	670	1459	2009	1446	109	5756

A list of the names of 5566 South Australians who died in World War I has been published by the South Australian Genealogy and Heraldry Society [67]. This number is a little smaller than the 5756 service records downloaded from the Roll of Honour, with the excess most probably due to the slight imprecision in the key-word search. In the spirit of Butler [64], the sampled conditional distribution for known age and known calendar year that is specified in Table 3.6.3 has been applied to the accepted total of 5566 to produce an estimated number of war deaths at each age for each calendar year. These estimated numbers are shown in Table 3.6.4.

Table 3.6.5 shows the entries of Table 3.6.4 re-arranged to give the estimated number of war deaths at each age for each of the generations from 1881 to 1900. Although the numbers displayed in this manner are not directly required for the calculations that follow in this section, Table 3.6.5 may have general historical interest.

Table 3.6.4: Estimated number of war deaths for South Australian males 1915-19

Age	Year of Death					
	1915	1916	1917	1918	1919	Total
16	0	3	2	0	0	5
17	0	6	5	0	0	11
18	10	40	27	6	0	83
19	42	95	99	63	2	301
20	60	111	142	103	3	419
21	44	105	155	82	3	389
22	39	126	208	131	3	507
23	31	126	188	149	3	497
24	40	108	171	108	6	433
25	36	84	166	103	3	392
26	27	90	137	79	0	333
27	36	45	126	90	8	305
28	24	52	94	81	10	261
29	19	60	76	78	0	233
30	19	53	74	65	6	217
31	13	39	68	57	0	177
32	15	34	65	68	3	185
33	16	31	60	44	3	154
34	11	24	26	47	2	110
35	11	16	42	24	3	96
36	6	23	27	26	3	85
37	8	15	16	21	2	62
38	3	18	19	31	0	71
39	5	15	19	11	0	50
40	3	19	11	11	0	44
41	2	13	13	6	0	34
42	3	3	10	6	2	24
43	0	6	5	6	2	19
44	2	3	6	2	0	13
45	5	10	6	3	0	24
46	2	3	5	5	0	15
47	0	2	0	2	0	4
48	0	0	3	0	0	3
49	0	0	2	2	0	4
50+	0	4	0	2	0	6
Total	532	1382	2073	1512	67	5566

Table 3.6.5: Estimated number of war deaths of South Australian males for generations 1881-1900

	Generation									
Age	**1881**	**1882**	**1883**	**1884**	**1885**	**1886**	**1887**	**1888**	**1889**	**1890**
25	0	0	0	0	0	0	0	0	0	36
26	0	0	0	0	0	0	0	0	27	90
27	0	0	0	0	0	0	0	36	45	126
28	0	0	0	0	0	0	24	52	94	81
29	0	0	0	0	0	19	60	76	78	0
30	0	0	0	0	19	53	74	65	6	0
31	0	0	0	13	39	68	57	0	0	0
32	0	0	15	34	65	68	3	0	0	0
33	0	16	31	60	44	3	0	0	0	0
34	11	24	26	47	2	0	0	0	0	0
35	16	42	24	3	0	0	0	0	0	0
36	27	26	3	0	0	0	0	0	0	0
37	21	2	0	0	0	0	0	0	0	0
Total	75	110	99	157	169	211	218	229	250	333
Average age	36.3	35.3	34.2	33.5	32.3	31.4	30.3	29.2	28.5	27.3

	Generation									
Age	**1891**	**1892**	**1893**	**1894**	**1895**	**1896**	**1897**	**1898**	**1899**	**1900**
18	0	0	0	0	0	0	10	40	27	6
19	0	0	0	0	0	42	95	99	63	2
20	0	0	0	0	60	111	142	103	3	0
21	0	0	0	44	105	155	82	3	0	0
22	0	0	39	126	208	131	3	0	0	0
23	0	31	126	188	149	3	0	0	0	0
24	40	108	171	108	6	0	0	0	0	0
25	84	166	103	3	0	0	0	0	0	0
26	137	79	0	0	0	0	0	0	0	0
27	90	8	0	0	0	0	0	0	0	0
28	10	0	0	0	0	0	0	0	0	0
Total	361	392	439	469	528	442	332	245	93	8
Average age	26.4	25.3	24.3	23.3	22.4	21.4	20.4	19.8	19.2	18.8

For each generation between 1881 and 1900, Table 3.6.6 shows the bootstrap estimate of generation expected life with standard error calculated by using linearly interpolated values for all q-values defining the generation lifetime distribution function. The estimate of generation expected life obtained in this manner is considered and designated “baseline” and is included in Table 3.6.2 in the column marked (1). Additionally, through a modification of the computer program << generation.f >> (*see* Section 3.4), the linearly interpolated estimates of the relevant set of q-values corresponding to the years 1915-19 (*see* Table 3.6.1) have been multiplied by factors of 5, 10 and 15 respectively; the estimates of the other q-values also determined by linear interpolation are unchanged. The three bootstrap expected values obtained under these conditions are again shown appropriately labeled in Table 3.6.6.

The computer program << generation.f >> has also been modified to suitably incorporate for each generation the relevant estimated numbers of civilian deaths (*see* Table 3.6.2) and war deaths (*see* Table 3.6.4). This modification to the program also calculates age-specific population sizes for the calendar years 1915-19 by successive yearly adjustment of the census count of 1911, allowing for both progressive aging of the 1911 population, and for civilian and war deaths. For each generation, estimates for the relevant set of q-values corresponding to the years 1915-19 (*see* Table 3.6.1) are calculated directly using these imputed age-specific number of deaths and population sizes. The estimates of the other q-values defining the generation lifetime distribution function are determined by linear interpolation. The bootstrap estimate of generation expected life calculated from this composite set of q-values is shown in Table 3.6.6 in the column marked (2).

For each generation, the estimate of generation expected life shown in column (1) can be cautiously interpreted as the expected length of life if World War I and the Influenza Pandemic had not occurred. It must be remembered, however, that any residual effect of these events on population mortality is present in the estimates of the q-values obtained from the mortality data collected in the years around the censuses of 1921 an onwards. The estimate of generation expected life shown in column (2), derived from the only known data for the number of civilian and war deaths for the years 1915-19, is presented as the most accurate measure of the combined effect of the events of this period.

The estimates of generation expected life shown in columns (1) and (2) are compared by using the ratio given in the last column of Table 3.6.6. This table suggests that the generation of males of 1895 was most affected, in a population sense, by World War I and the Influenza Pandemic with a reduction of approximately 8% in expected life from the baseline level. By comparison with the estimates of generation expected life calculated by using the factor multiples of the baseline q-values, the expected value of 52.75 years for the generation of 1895 is equivalent to the estimate derived with an average ten-fold increase in the baseline mortality rate over the years 1915-19. This conclusion is consistent with the estimated number of war deaths for the generation of 1895 shown in Table 3.6.5, and the average number of civilian deaths for ages 20-25 for 1915-19 shown in Table 3.6.2.

A different insight into the effects of World War I on generation expected life is obtained by considering each generation as a mixture of two sub-populations. One sub-population consists of those individuals who left, or using the official terminology "embarked from", Australia on overseas military service. The other sub-population consists of the remainder of the generation; those who did not enlist in the armed forces, and those who did enlist but did not embark. The generation expected life of this latter group is assumed to be approximately GEL(1), since while a "baseline" estimate does not include the effects of World War I, it also does not include the effects of the additional mortality produced by the Influenza Pandemic. However this slight deficiency is not considered to be of any great consequence for the order argument presented here. Denoting the generation expected life of the "embarkation" sub-population as GEL(3), then

$$\text{GEL}(2) = p\ \text{GEL}(3) + (1-p)\ \text{GEL}(1)$$

where p is the proportion of "embarkations" in the generation, from which it follows that

$$\text{GEL}(3) = [\ \text{GEL}(2) - (1 - p)\ \text{GEL}(1)\ ] / p$$

Hence an estimate of GEL(3) can be determined for each generation if an estimate of p is available, since estimates of GEL(1) and GEL(2) are given in Table 3.6.6.

Table 3.6.6: Comparison of estimates of generation expected life of South Australian males 1881-1900

Generation	GEL from Baseline data (1)	GEL using factor multiplier of Baseline q-values from 1915-19			GEL from 1915-19 data (2)	Ratio (%) GEL(2):GEL(1)
		Factor				
		5	10	15		
1881	51.71 (.40)	49.10	46.15	43.53	51.05 (.39)	98.7
1882	52.25 (.38)	49.66	46.73	44.12	51.37 (.36)	98.3
1883	52.80 (.36)	50.17	47.18	44.48	52.04 (.34)	98.6
1884	53.36 (.34)	50.78	47.83	45.15	52.20 (.32)	97.8
1885	53.92 (.32)	51.41	48.52	45.88	52.65 (.31)	97.6
1886	54.46 (.32)	52.06	49.26	46.71	52.87 (.30)	97.1
1887	55.00 (.32)	52.71	50.04	47.56	53.31 (.30)	96.9
1888	55.52 (.33)	53.23	50.55	48.07	53.72 (.31)	96.8
1889	56.02 (.34)	53.72	51.02	48.53	54.04 (.32)	96.5
1890	56.53 (.36)	54.16	51.39	48.78	53.87 (.34)	95.3
1891	56.97 (.37)	54.55	51.73	49.08	54.03 (.34)	94.8
1892	57.08 (.35)	54.78	52.04	49.48	53.81 (.31)	94.3
1893	57.22 (.33)	54.87	52.09	49.51	53.47 (.30)	93.4
1894	57.37 (.32)	55.02	52.23	49.63	53.21 (.28)	92.7
1895	57.50 (.31)	55.12	52.31	49.64	52.75 (.28)	91.7
1896	57.63 (.31)	55.20	52.33	49.65	53.45 (.29)	92.7
1897	57.76 (.33)	55.30	52.42	49.68	54.45 (.29)	94.3
1898	57.91 (.33)	55.94	53.61	51.38	55.41 (.32)	95.7
1899	58.07 (.35)	56.71	55.06	53.45	57.07 (.34)	98.3
1900	58.27 (.37)	57.31	56.10	54.93	58.12 (.37)	99.7

Data is not available to immediately and directly estimate p for any generation of South Australian males. The official record of the 1/AIF [63] states that there were 27,761 embarkations from South Australia of the 34,566 enlistments over the period 1914-18. Butler [64] gives the number of national embarkations for each year of the 1/AIF, and these figures, expressed as proportions, are shown in Table 3.6.7. The Australian distribution has been applied to the total number of embarkations from South Australia to produce the estimated number of yearly embarkations from South Australia that is shown in Table 3.6.7. Butler also gives a national distribution of age at embarkation of the 1/AIF, also shown in Table 3.6.7. This national distribution has been applied to each of the estimated number of yearly embarkations from South Australia. The results of these calculations are shown as the cell entries in Table 3.6.7, with each corresponding generation indicated in parenthesis. For example, an 18 year old embarking in 1915, a 19 year old embarking in 1916 and a 20 year old embarking in 1917 were all born in 1897 and hence, by definition, belong to the same generation. Embarkations of individuals of the same generation are alternatively indicated by the shading pattern on the cells of Table 3.6.7, reading from the left and down diagonally.

For each generation, Table 3.6.8 shows the number of embarkations obtained after appropriate accumulation, and the age and population size of the generation in 1915, from which the proportion p has been calculated. The estimate of p has been combined with GEL(1) and GEL(2) in the manner previously described to produce GEL(3), the generation expected life of the sub-population of the generation who embarked for overseas military service during World War I. This value is given in Table 3.6.9, where the ratio of GEL(3) to GEL(1) is also shown. The consequences of World War I were greatest for the sub-population with overseas military service of the generations of 1890 to 1895, with generation expected life approximately 85% of the generation expected life of the sub-population who remained in South Australia.

Despite the assumptions made and approximations used , the estimates of generation expected life presented in this section are considered to be reasonable measures of the order of the effect of the events of 1914-19 on the mortality of South Australian males for the generations of 1881-1900.

Table 3.6.7: Estimates of the number of embarkations to overseas military service of South Australian males 1915-18

		1915	**1916**	**1917**	**1918**
	Australian embarkations	.342	.414	.174	.070
		Estimated number of yearly embarkations from South Australia			
		1915	**1916**	**1917**	**1918**
		9495	11493	4830	1943
		Estimated number of age-specific yearly embarkations from South Australia and (generation)			
Age		**1915**	**1916**	**1917**	**1918**
18	.0708	672 (1897)	814 (1898)	342 (1899)	138 (1900)
19	.0769	730 (1896)	884 (1897)	371 (1898)	149 (1899)
20	.0830	788 (1895)	955 (1896)	402 (1897)	161 (1898)
21	.0892	846 (1894)	1026 (1895)	431 (1896)	172 (1897)
22	.0778	739 (1893)	894 (1894)	376 (1895)	151 (1896)
23	.0673	639 (1892)	773 (1893)	325 (1894)	131 (1895)
24	.0603	573 (1891)	693 (1892)	291 (1893)	117 (1894)
25	.0533	506 (1890)	613 (1891)	257 (1892)	104 (1893)
26	.0472	448 (1889)	542 (1890)	228 (1891)	92 (1892)
27	.0425	404 (1888)	488 (1889)	205 (1890)	83 (1891)
28	.0385	366 (1887)	442 (1888)	186 (1889)	75 (1890)
29	.0350	332 (1886)	402 (1887)	169 (1888)	68 (1889)
30	.0315	299 (1885)	362 (1886)	152 (1887)	61 (1888)

Table 3.6.7: (continued)

Age		1915	1916	1917	1918
31	.0284	270 (1884)	326 (1885)	137 (1886)	55 (1887)
32	.0258	245 (1883)	297 (1884)	125 (1885)	50 (1886)
33	.0233	221 (1882)	268 (1883)	113 (1884)	45 (1885)
34	.0213	202 (1881)	245 (1882)	103 (1873)	41 (1884)
35	.0192	182 (1880)	221 (1881)	93 (1882)	37 (1883)
36	.0173	164 (1879)	199 (1880)	84 (1881)	34 (1882)
37	.0154	146 (1878)	177 (1879)	74 (1880)	30 (1881)
38	.0137	130 (1877)	157 (1878)	66 (1879)	27 (1880)
39	.0120	114 (1876)	138 (1877)	58 (1878)	23 (1879)
40	.0108	103 (1875)	124 (1876)	52 (1877)	21 (1878)
41	.0096	91 (1874)	110 (1875)	46 (1876)	19 (1877)
42	.0087	83 (1873)	100 (1874)	42 (1875)	17 (1876)
43	.0087	83 (1872)	100 (1873)	42 (1874)	17 (1875)
44	.0096	91 (1871)	110 (1872)	46 (1873)	19 (1874)
45	.0029	28 (1870)	33 (1871)	14 (1872)	6 (1873)

Table 3.6.8: Estimates of the proportion of South Australian males with overseas military service for the generations of 1881-97

Generation	(Age) and Population in 1915	Embarkations for 1915-19	p
1881	(34) 3300	537	.163
1882	(33) 3474	593	.171
1883	(32) 3654	653	.179
1884	(31) 3840	721	.188
1885	(30) 4020	795	.198
1886	(29) 4158	881	.212
1887	(28) 4260	975	.229
1888	(27) 4332	1076	.248
1889	(26) 4368	1190	.272
1890	(25) 4398	1328	.302
1891	(24) 4416	1497	.339
1892	(23) 4416	1681	.381
1893	(22) 4398	1907	.434
1894	(21) 4350	2182	.502
1895	(20) 4272	2321	.543
1896	(19) 4170	2267	.544
1897	(18) 4056	2130	.525

Table 3.6.9: Estimates of generation expected life of South Australian males with overseas military service

Generation	p	GEL from Baseline data (1)	GEL from 1915-19 data (2)	GEL for Embarkations (3)	Ratio (%) GEL(3):GEL(1)
1881	.163	51.71	51.05	47.66	92.2
1882	.171	52.25	51.37	47.10	90.2
1883	.179	52.80	52.04	48.55	92.0
1884	.188	53.36	52.20	47.19	88.4
1885	.198	53.92	52.65	47.51	88.1
1886	.212	54.46	52.87	46.96	86.2
1887	.229	55.00	53.31	47.62	86.6
1888	.248	55.52	53.72	48.26	86.9
1889	.272	56.02	54.04	48.74	87.0
1890	.302	56.53	53.87	47.72	84.4
1891	.339	56.97	54.03	48.30	84.8
1892	.381	57.08	53.81	48.50	85.0
1893	.434	57.22	53.47	48.58	84.9
1894	.502	57.37	53.21	49.08	85.6
1895	.543	57.50	52.75	48.75	84.8
1896	.544	57.63	53.45	49.95	86.7
1897	.525	57.76	54.45	51.46	89.1

SUMMARY AND CONCLUSIONS

The principal results of this thesis are contained in Table 2.16.1 (page 131), where the estimates of current expected life for the years 1841-1996 are shown for each gender. These estimates are also displayed graphically in Figure 2.16.1 (page 134). The majority of these estimates have not been calculated prior to this thesis. Over the period 1841-1996, the average length of life in South Australia, as measured by the statistic current expected life, has nearly doubled for both genders. The estimates for 1841, 1844 and 1846 are based on both limited and imputed mortality and population data, and have been included for completeness and their historical curiosity. Nevertheless, for each gender, the reasonable consistency of the estimates for these three years with the estimate for 1851 illustrates a robust quality of the general life table methodology.

Bootstrap estimates of the standard error of the estimate of current expected life are also shown in Table 2.16.1, and these estimates are graphed in Figure 2.16.2 (page 135) against the square root of the total population size on which each estimate is based. An empirical linear prediction formula for the standard error has been derived from the South Australian estimates. Table 2.16.1 and Figure 2.16.2 suggests that this formula is appropriate for similar populations ranging in size from 20,000 to 700,000 approximately. It is clear from Table 2.16.1 that the standard error of the estimate of current expected life is not negligible, even in large populations. The South Australian population of 699,000 males in 1996 has an estimated current expected life of 75.33 years, with a bootstrap standard error of .17 years. The investigations in this thesis into the sources of variation in the estimate of current expected life indicate that the variation is predominantly a consequence of random variation in the numbers of deaths rather than random variation in population sizes. The relevance of standard errors depends on circumstances. In this thesis, standard errors have been used to indicate the precision of the estimates of current expected life, and they have also been used in the calculation of asymptotic z-scores. From these z-scores, conclusions have been made about numerical differences in gender estimates of current expected life for a given year, and about numerical differences in thesis estimates of current expected life and comparable estimates from external sources. From this perspective, given the results in this thesis, it is difficult to know how much credence to place on the unsupported statement by Weir [68], where it is stated that “average life expectancy was around forty years” in Elizabethan England. Furthermore, with the range in size of the standard errors estimated in this thesis, it is unclear how a modification of life table methodology proposed by Pollard [69] that produces a change in the estimated value of current expected life of “two hundredths of a year of life greater”, “could be important in a comparative situation”.

Within each section of Chapter 2, four estimates of current expected life have been calculated for each gender. These estimates result from the combination of each technical method of calculation (q-method & u-method; *see* Section 1.7 and Section 1.8) with each type of life table (complete, with mortality and population data tabulated using one-year age groupings & abridged, with mortality and population data tabulated using a mixture of one-year and five-year age groupings; *see* Section 1.7 and Section 1.8). For current notational convenience, the four types of estimate are designated qc, qa, uc and ua. Overall, there is essentially no difference between type qc, type qa and type uc in the applications considered in this thesis, with estimates of the type ua generally being slightly larger than the other three types in any particular situation. In most instances these numerical differences are statistically insignificant when standard errors are considered. Therefore, by virtue of computational simplicity, estimates of type qc and type qa that require only simple arithmetic operations are considered preferable to estimates of type uc and type ua that require both arithmetical and exponential operations. The comparison of estimates of type qc and type qa has the most immediate relevance for practical purposes, since a type qc estimate requires more detailed mortality and population distributions than does a type qa estimate. It was possible to calculate twenty (10 censuses from the 20th century x 2 genders) estimates of type qc for this thesis. Corresponding estimates of type qa have been calculated from the same data after a simple accumulation of the one-year data using the appropriate mixture of one-year and five-year age levels. The twenty pairs of estimates are “perfectly” correlated with a Pearson correlation coefficient of 1.000, and have a maximum value of 100.063% for the ratio of the estimate values type qa / type qc. From these examples, there appears to be no essential difference between a type qc estimate of current expected life and the corresponding type qa estimate of current expected life. There thus appears to be no advantage, for practical purposes, in collecting and tabulating data in the one hundred single-year age groupings needed for a type qc estimate of current expected life. This observation is in agreement with the conclusions reached by Doering [51] (*see* Section 2.11) and his “skeleton” life table based on fewer and broader age groupings.

Within each section of Chapter 2, a robustness examination has been carried out for each gender. Three possibilities are compared to the standard estimate (type qc or type qa, as appropriate) of current expected life showing the effects of proportionally increasing the recorded numbers of deaths or the recorded population sizes. It is generally accepted that population sizes determined from a census are less accurately measured than are yearly numbers of deaths registered because of a civil legal obligation. Consequently a standard estimate of current expected life based on recorded numbers of deaths and recorded population sizes is a biased underestimate. The potential extent of the bias under a range of circumstances is illustrated by estimates obtained after adjusting the recorded population sizes by a “best guess” (upwards by either 2% or 5%) and by a “worst case” (upwards by 10%). For some calendar years, the numerical differences between the various estimates are six times the standard errors of the estimates. However, while the intent of the robustness investigations is to indicate the approximate effects of variation in assumptions, it must be remembered that the value of the estimate of the fundamental statistic ${}_{j}q_{k}$ is unchanged by an equal proportional increase or decrease in the number of deaths and the population size.

In Chapter 3 the generation life table method of measuring the average lifetime of a population has been applied to notional populations born in various calendar years. Although estimates of average lifetime obtained by this method are retrospective, they are the closest measures of “true” average lifetimes obtainable from routinely collected data. In this sense, generation expected life can be viewed as the “correct answer” against which the “prediction” made by current expected life for a given year can be assessed. The examples presented in Chapter 3 for the years 1851, 1881 and 1901show some of the vagaries encountered in using current expected life as a predictor of future average lifetime. For 1901, the outcome is that suggested in Section 1.7. The generally regularly improving social and living conditions in South Australia over the 20th Century has resulted in current expected life being less than generation expected life, for both males and females. Conversely, the adverse conditions subsequent to 1851 and which are reflected in the pattern of the estimates of current expected life for 1855-81, have resulted in the estimates of current expected life for 1851 to be approximately equal to the corresponding estimates of generation expected life for 1851. The examples given in Section 3.6 of generation expected life for 1881-99 illustrate that the generation life table methodology combined with appropriate data is "essential to a correct historical study of ….population growth" (Lancaster [59]). The estimates of generation expected life for 1890-95, particularly, shown in Table 3.6.9 are stark numerical measures of "the broken years" of Gammage [66].

****

> "We are all beset by statistics. It is a cultural trait of modern western man to collect and consume numbers, and more statistics have been collected and published about human populations than about anything else. The snare lies not so much with the huge deposit of statistical ore but in its refinement."
>
> Webb (Chapter 9, [70])

APPENDIX: CD-ROM OF COMPUTER PROGRAMS, DATA FILES AND OUTPUT TABLES

The computer files listed below are included on the accompanying CD-rom. Details of these files are listed in Sections 2.3 - 2.6 and Section 3.4.

Also included on the CD-rom is the post-script file << thesis.ps >>, created by using the "print-to-file" command in Word 2000 for an Apple LaserWriter 12/640 PS.

Program files	Data files		Output (result) files: q-method		Output (result) files: u-method	
current.f	M1841.txt	F1841.txt	CELQM1841.txt	CELQF1841.txt	CELUM1841.txt	CELUF1841.txt
current.exe	M1844.txt	F1844.txt	CELQM1844.txt	CELQF1844.txt	CELUM1844.txt	CELUF1844.txt
currentin.txt	M1846.txt	F1846.txt	CELQM1846.txt	CELQF1846.txt	CELUM1846.txt	CELUF1846.txt
bootstrap.txt	M1851.txt	F1851.txt	CELQM1851.txt	CELQF1851.txt	CELUM1851.txt	CELUF1851.txt
	M1855.txt	F1855.txt	CELQM1855.txt	CELQF1855.txt	CELUM1855.txt	CELUF1855.txt
generation.f	M1861.txt	F1861.txt	CELQM1861.txt	CELQF1861.txt	CELUM1861.txt	CELUF1861.txt
generation.exe	M1866.txt	F1866.txt	CELQM1866.txt	CELQF1866.txt	CELUM1866.txt	CELUF1866.txt
generationin.txt	M1871.txt	F1871.txt	CELQM1871.txt	CELQF1871.txt	CELUM1871.txt	CELUF1871.txt
generationout.txt	M1876.txt	F1876.txt	CELQM1876.txt	CELQF1876.txt	CELUM1876.txt	CELUF1876.txt
	M1881.txt	F1881.txt	CELQM1881.txt	CELQF1881.txt	CELUM1881.txt	CELUF1881.txt
thesis.ps	M1891.txt	F1891.txt	CELQM1891.txt	CELQF1891.txt	CELUM1891.txt	CELUF1891.txt
	M1901.txt	F1901.txt	CELQM1901.txt	CELQF1901.txt	CELUM1901.txt	CELUF1901.txt
	M1911.txt	F1911.txt	CELQM1911.txt	CELQF1911.txt	CELUM1911.txt	CELUF1911.txt
	M1921.txt	F1921.txt	CELQM1921.txt	CELQF1921.txt	CELUM1921.txt	CELUF1921.txt
	M1933.txt	F1933.txt	CELQM1933.txt	CELQF1933.txt	CELUM1933.txt	CELUF1933.txt
	M1947.txt	F1947.txt	CELQM1947.txt	CELQF1947.txt	CELUM1947.txt	CELUF1947.txt
	M1954.txt	F1954.txt	CELQM1954.txt	CELQF1954.txt	CELUM1954.txt	CELUF1954.txt
	M1961.txt	F1961.txt	CELQM1961.txt	CELQF1961.txt	CELUM1961.txt	CELUF1961.txt
	M1966.txt	F1966.txt	CELQM1966.txt	CELQF1966.txt	CELUM1966.txt	CELUF1966.txt
	M1971.txt	F1971.txt	CELQM1971.txt	CELQF1971.txt	CELUM1971.txt	CELUF1971.txt
	M1976.txt	F1976.txt	CELQM1976.txt	CELQF1976.txt	CELUM1976.txt	CELUF1976.txt
	M1981.txt	F1981.txt	CELQM1981.txt	CELQF1981.txt	CELUM1981.txt	CELUF1981.txt
	M1986.txt	F1986.txt	CELQM1986.txt	CELQF1986.txt	CELUM1986.txt	CELUF1986.txt
	M1991.txt	F1991.txt	CELQM1991.txt	CELQF1991.txt	CELUM1991.txt	CELUF1991.txt
	M1996.txt	F1996.txt	CELQM1996.txt	CELQF1996.txt	CELUM1996.txt	CELUF1996.txt

This thesis has been written using Microsoft Word 2000, mathematical equation editor MathType [71], and bibliography reference manager EndNote [72].

REFERENCES

1. G.M. Tallis & P. Leppard : *Is length of life predictable ?*
   Human Biology **69** (6) pp 873-886. (1997)
2. M. Spiegelman : *Introduction to Demography*
   New York, The Society of Actuaries. (1955)
3. Anonymous : *Celebrating 113 years* Adelaide, The Advertiser. (2001 March)
4. Anonymous : *Oldest person dies* Adelaide, The Advertiser. (2002 March)
5. Anonymous : *Australia's oldest person dies after 114 happy years*
   Adelaide, The Advertiser. (2002 May)
6. J. Graunt : *Natural and political observations made upon the bills of mortality (reprinted 1939)* Baltimore, Johns Hopkins Press. (1662)
7. E. Halley : *An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslau*
   Philosophical transactions of the Royal Society **17** pp 596-610. (1693)
8. R.C. Elandt-Johnson & N.L. Johnson : *Survival Models and Data Analysis*
   New York, John Wiley & Sons. (1980)
9. D.R. Cox & D. Oakes : *Analysis of Survival Data* London, Chapman & Hall. (1984)
10. B. Benjamin & H.W. Haycocks : *The Analysis of Mortality and other Actuarial Statistics* Cambridge, Cambridge University Press. (1970)
11. S. Selvin : *Statistical Analysis of Epidemiologic Data*
    Oxford, Oxford University Press. (1996)
12. C. Newell : *Methods and Models in Demography* London, Belhaven Press. (1988)
13. K. Namboodiri & C.M. Suchindran : *Life Table Techniques and their Applications*
    New York, Academic Press. (1987)
14. C.L. Chiang : *The Life Table and its Applications*
    Malabar, Florida, Robert E. Krieger Publishing Company. (1984)
15. Anonymous : *Longer life nears limit* Adelaide, The Advertiser. (2001 February)
16. W. Feller : *An Introduction to Probability Theory and its Applications*
    New York, John Wiley & Sons. (1966)

17. M. Abramowitz & I.A. Stegun : *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*
Washington, U.S. Government Printing Office. (1965)

18. M.R. Spiegel : *Mathematical Handbook of Formulas and Tables*
New York, McGraw-Hill Book Company. (1968)

19. P.R. Cox : *Demography* Cambridge, Cambridge University Press. (1976)

20. N.L. Johnson, S. Kotz & N. Balakrishnan : *Continuous Univariate Distributions*
New York, John Wiley & Sons. (1994)

21. C.L. Chiang : *A stochastic study of the life table and its applications: I Probability distributions of the biometric functions*
Biometrics **16** (December) pp 618-635. (1960)

22. C.L. Chiang : *A stochastic study of the life table and its applications: II Sample variance of the observed expectation of life and other biometric functions*
Human Biology **32** (3) pp 221-238. (1960)

23. C.L. Chiang : *On constructing current life tables*
Journal of the American Statistical Association **67** (339) pp 538-541. (1972)

24. A.M. Mood & F.A. Graybill : *Introduction to the theory of statistics*
New York, McGraw-Hill Book Company. (1963)

25. M.S. Kendall, A. Stuart & K. Ord : *Kendall's Advanced Theory of Statistics*
London, Charles Griffin & Company. (1987)

26. S. Kotz & N.L. Johnson : *Encyclopedia of Statistical Sciences*
New York, John Wiley & Sons. (1982)

27. *IMSL FORTRAN subroutines for statistical analysis* Houston, IMSL (Inc). (1991)

28. B. Efron : *Bootstrap methods: another look at the jacknife*
Annals of Statistics **7** pp 1-26. (1979)

29. B. Efron & R.J. Tibshirani : *An Introduction to the Bootstrap*
New York, Chapman & Hall. (1993)

30. W.J. Kennedy & J.E. Gentle : *Statistical Computing*
New York, Marcel Dekker. (1980)

31. B.A. Wichmann & I.D. Hill : *An efficient and portable pseudo-random number generator* Applied Statistics **31** pp 188-190. (1982)

32. J.E. Gentle : *Random number generation and Monte Carlo methods*
New York, Springer-Verlag. (1998)

33. V. Chew : *Some useful alternatives to the normal distribution*
The American Statistician **22** (3) pp 22-24. (1968)

34. N.L. Johnson, S. Kotz & A.W. Kemp : *Univariate discrete distributions*
New York, John Wiley & Sons. (1992)

35. J. Camm : *The early nineteenth century colonial censuses of Australia*
Sydney, Project Australia 1788-1988 : A Bicentennial History
(History Project Incorporated). (1988)

36. T. Stevenson : *Australasian census development with special reference to colonial South Australia : a resource paper*
Sydney, Project Australia 1788-1988 : A Bicentennial History
(History Project Incorporated). (1981)

37. A.E. Miller : *Checklist of Nineteenth Century Australian Colonial Statistical Sources: Censuses, Blue Books and Statistical Registers*
Sydney, Project Australia 1788-1988 : A Bicentennial History
(History Project Incorporated). (1982)

38. A.G. Peake : *Sources for South Australian History*
Adelaide, Tudor Australia Press. (1987)

39. M. Pitt : *Official statistics in Australia : the "Blue Books" and the development of pre-contemporary official statistics* Canberra, Australian Bureau of Statistics. (1986)

40. G. Pennock : *Foundation to Federation : South Australian statistical sources 1836-1901* Adelaide, Australian Bureau of Statistics. (1990)

41. I. Castles : *[Microfiche] Catalogue of Australian statistical publications 1804 to 1901*
Canberra, Australian Bureau of Statistics. (1989)

42. *Pro Fortran for Windows* Rochester Hills, Michigan, Absoft Corporation. (1999)

43. C.H. Wickens : *Australian mortality*
Journal of the Institute of Actuaries **61** pp 165-213. (1930)

44. M.B. Pell : *On the rates of mortality and expectation of life in New South Wales as compared with England and other countries*
Transactions of the Royal Society of New South Wales **1** pp 66-76. (1867)

45. M.B. Pell : *On the rates of mortality in New South Wales, and on the construction of mortality tables from census returns; with a note on the formation of commutation tables* Journal of the Institute of Actuaries **21** pp 257-288. (1879)

46. A.F. Burridge : *On the rates of mortality in Australia*
Journal of the Institute of Actuaries **24** pp 333-358. (1884)

47. G.W. Snedecor & W.G. Cochran : *Statistical methods*
Ames, Iowa, Iowa State University Press. (1967)

48. C. Thorburn : *Australian life tables 1993-95*
Canberra, Australian Government Actuary. (1997)

49. P.C. Wickens : *Abridged life tables for the Australian States, 1953-1955*
The Australian Journal of Statistics **2** (1) pp 114 - 121. (1960)

50. C.M. Young : *Population of Australia (Chapter VII - Mortality)*
Bancock, United Nations Economic and Social Commission for Asia and the Pacific. (1982)

51. C.R. Doering & A.L. Forbes : *A skeleton life table* Proceedings of the National Academy of Sciences of the United States of America **24** pp 400-405. (1938)

52. I.W. Burr : *Cumulative frequency functions*
The Annals of Mathematical Statistics **13** pp 215-232. (1942)

53. R.N. Rodriguez : *A guide to Burr type XII distributions*
Biometrika **64** (1) pp 129-134. (1977)

54. W. Dixon : *BMDP Statistical Software* Los Angeles, UCLA Press. (1993)

55. *Twelfth annual report of the Registrar-General of Births, Deaths and Marriages in England* London, Her Majesty's Printer. (1853)

56. P. Woodruff : *Two million South Australians*
Kent Town, South Australia, Peacock Publications. (1984)

57. I.L. Forbes : *From Colonial Surgeon to Health Commission - The Government provision of health services in South Australia 1836-1995*
Adelaide, Openbook. (1996)

58. M. Greenwood : *Notes on the New National Life Tables ..discussion (J.M.Laing)*
Journal of the Institute of Actuaries **59** pp 215-216. (1928)

59. H.O. Lancaster : *Generation life tables for Australia*
The Australian Journal of Statistics **1** (1) pp 19-33. (1959)

60. C.M. Young : *An analysis of the population growth and mortality of selected birth cohorts in Australia, with reference to the relationship between cohort and transverse (or calendar year) experience* Canberra, Australian National University. (1969)

61. K.D. Patterson & G.F. Pyle : *The geography and mortality of the 1918 influenza pandemic* Bulletin of the History of Medicine **65** pp 4-21. (1991)

62. C.E.W. Bean : *The official history of Australia in the war of 1914-1918* St. Lucia, Queensland, University of Queensland Press. (1981)

63. *Australian Imperial Force : statistics of casualties, etc. compiled to 30th June, 1919* London, The British Army Records Section. (1919)

64. A.G. Butler : *The Australian Army Medical Services in the War of 1914-1918* Canberra, Australian War Memorial. (1943)

65. C.B. Lock : *The Fighting 10th (10th Battalion, Australian Imperial Force, 1914-19)* Adelaide, Webb & Son. (1936)

66. W. Gammage : *The Broken Years* Canberra, Australian National University Press. (1974)

67. *South Australian casualties of World War I : the Great War roll of honour* Adelaide, South Australian Genealogy and Heraldry Society. (1994)

68. A. Weir : *Elizabeth the Queen* London, Random House. (1998)

69. J.H. Pollard : *On the derivation of a full life table from mortality data recorded in five-year age groups* Mathematical Population Studies **2** (1) pp 1-14. (1989)

70. R.U. Cooke & J.H. Johnson : *Trends in Geography* Oxford, Pergamon Press. (1969)

71. *MathType* Long Beach, California, Design Science. (1999)

72. *EndNote* Berkeley, ISI ResearchSoft. (1998)