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 iii
DECLARATION iv
LIST OF TABLES v
LIST OF FIGURES viii
ACKNOWLEDGEMENTS ix

INTRODUCTION 1

CHAPTER 1: CURRENT LIFE TABLE ANALYSIS AND CURRENT EXPECTED LIFE 5
1.1 Introduction 5
1.2 Current life table analysis 7
1.3 Derivation of current life table analysis 8
1.3.1 A distribution function expressed as a product of conditional probabilities 9
1.3.2 A distribution function specified by a hazard function 9
1.3.3 Using the distribution function to derive the expected value of a random variable 10
1.4 Mathematical approximations 10
1.4.1 Trapezoidal rule for numerical integration 10
1.4.2 Mean value rule for numerical integration 10
1.4.3 Linear interpolation 10
1.5 Age-specific expected number of deaths and expected size of the population 11
1.6 Some comments about the fundamental estimator 21
1.7 Definition of current life table analyses 22
1.8 An alternative approach using the hazard estimator 25
1.9 Estimation of lifetime distribution functions for current life table analysis 26
1.10 Estimation of current expected life 29
1.11 Estimation of the variance of the estimate of current expected life 30
1.11.1 The Triangular Distribution: T(a,m,b) 32
1.11.2 Approximation of the Standard Normal distribution by the Triangular 34
1.11.3 Bootstrap estimation of the variance of the estimate of current expected life 35
1.11.4 An additional source of variation in the estimate of current expected life 36
1.11.5 A robustness determination of the estimate of current expected life 37
1.12 A comparison of methodologies 38

CHAPTER 2: APPLICATION TO SOUTH AUSTRALIAN DATA 41
2.1 The sources of data for the estimation of current expected life for South Australia 41
2.2 The computing environment of the computer programs used for this thesis 44
2.3 The naming and structure of data files 44
2.4 The thesis FORTRAN computer program for current expected life 45
2.5 The specification of bootstrap analyses 46
2.6 Current life table analyses for the period 1971-1996 51
2.7 Current life table analyses for 1961 and 1966 62
2.8 Current life table analyses for the period 1933-1954 65
2.9 Current life table analyses for 1911 and 1921 69
2.10 Current life table analyses for the period 1876-1901 73
2.11 Current life table analyses for 1871 79
2.12 Current life table analyses for 1861 and 1866 89
2.13 Current life table analyses for 1851 and 1855 96
2.14 Current life table analyses for 1844 and 1846 110
2.15 Current life table analyses for 1841 123
2.16 Summary and discussion 131

CHAPTER 3: GENERATION EXPECTED LIFE FOR SOUTH AUSTRALIA 136
3.1 Introduction 136
3.2 Rationale for a generation lifetime distribution function 139
3.3 Estimation of generation expected life 141
3.4 The thesis FORTRAN computer program for generation expected life 142
3.5 Some estimates of generation expected life for South Australia 143
3.6 The effect of the events of 1914-19 on male generation expected life 1881-1900 145

SUMMARY AND CONCLUSIONS 161

APPENDIX: Index to CD-rom of computer files 165

REFERENCES 166

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

Table 1.12 Comparison of Chiang and Thesis results for US White Males 1955 39

Table 2.3.1 An extract from a typical data file.....45
Table 2.5.1 Estimates of error rates in the Censuses of 1911 and 1921.....47
Table 2.5.2 Parameter values selected for bootstrap analyses 50
Table 2.6.1 Sources of population and mortality data for the period 1971-1996 51
Table 2.6.2 An example of a typical output file.....53
Table 2.6.3 Estimates of current expected life for the period 1971-1996.....55
Table 2.6.4 Comparison of estimates of current expected life for the period 1971-1996....56
Table 2.6.5 Methodological comparisons for the period 1971-1996 58
Table 2.6.6 Robustness comparisons for the period 1971-1996 61
Table 2.7.1 Sources of population and mortality data for 1961 and 1966 62
Table 2.7.2 Estimates of current expected life for 1961 and 1966.....63
Table 2.7.3 Methodological comparisons for 1961 and 1966.....63
Table 2.7.4 Robustness comparisons for 1961 and 1966.....64
Table 2.8.1 Sources of population and mortality data for the period 1933-1954 65
Table 2.8.2 Estimates of current expected life for the period 1933-1954.....66
Table 2.8.3 Methodological comparisons for the period 1933-1954 67
Table 2.8.4 Robustness comparisons for the period 1933-1954 68
Table 2.9.1 Sources of population and mortality data for 1911 and 1921 69
Table 2.9.2 Estimates of current expected life for 1911 and 1921.....70
Table 2.9.3 Methodological comparisons for 1911 and 1921.....71
Table 2.9.4 Robustness comparisons for 1911 and 1921 71
Table 2.10.1 Sources of population and mortality data for the period 1876-1901 73
Table 2.10.2 Estimates of current expected life for the period 1876-1901 74
Table 2.10.3 Comparison of estimates of current expected life for the period 1876-1901....75
Table 2.10.4 Methodological comparisons for the period 1876-1901 75
Table 2.10.5 Robustness comparisons for the period 1876-1901 77
Table 2.11.1 Sources of population and mortality data for 1871 79
Table 2.11.2 Estimated population sizes for 1871: Ages 80+.....80

Table 2.11.3 Comparison of the number of deaths from two sources for 1870-72 81
Table 2.11.4 Comparison of sampled and estimated number of deaths for 1870-72.....83
Table 2.11.5 Estimates of current expected life for 1871 85
Table 2.11.6 Methodological comparisons for 1871.....86
Table 2.11.7 Robustness comparisons for 1871.....87
Table 2.12.1 Sources of population and mortality data for 1861 and 1866 89
Table 2.12.2 Estimated population sizes for 1861 and 1866: Ages 80+.....90
Table 2.12.3 Comparison of the number of deaths from two sources for 1861.....91
Table 2.12.4 Comparison of sampled and estimated number of deaths for 1861 92
Table 2.12.5 Estimates of current expected life for 1861 and 1866.....93
Table 2.12.6 Methodological comparisons for 1861 and 1866.....94
Table 2.12.7 Robustness comparisons for 1861 and 1866.....95
Table 2.13.1 Sources of population and mortality data for 1851 and 1855 96
Table 2.13.2 Estimated population sizes for 1851 and 1855: Ages 0-2.....97
Table 2.13.3 Census population counts for 1851 and 1855: Ages 2-60.....98
Table 2.13.4 Estimated population sizes for 1851 and 1855: Ages 60+.....100
Table 2.13.5 Comparison of the number of deaths from two sources for 1850-52 101
Table 2.13.6 Comparison of the number of deaths from two sources for 1854-55 102
Table 2.13.7 Comparison of sampled and estimated number of deaths for 1850-52.....103
Table 2.13.8 Comparison of sampled and estimated number of deaths for 1854-55.....105
Table 2.13.9 Estimates of current expected life for 1851 and 1855.....107
Table 2.13.10 Methodological comparisons for 1851 and 1855.....108
Table 2.13.11 Robustness comparisons for 1851 and 1855.....109
Table 2.14.1 Sources of population and mortality data for 1844 and 1846 110
Table 2.14.2 Estimated population sizes for 1844 and 1846: Ages 0-2.....112
Table 2.14.3 Census population counts for 1844 and 1846: Ages 2-60.....112
Table 2.14.4 Estimated population sizes for 1844 and 1846: Ages 60+.....113
Table 2.14.5 Comparison of the number of deaths from two sources for 1843-45 114
Table 2.14.6 Comparison of the number of deaths from two sources for 1845-47 114
Table 2.14.7 Estimated total population sizes and total number of deaths for 1841-45 115
Table 2.14.8 Comparison of sampled and estimated number of deaths for 1843-45.....117
Table 2.14.9 Comparison of sampled and estimated number of deaths for 1845-47.....119

Table 2.14.10 Estimates of current expected life for 1844 and 1846 121
Table 2.14.11 Methodological comparisons for 1844 and 1846 122
Table 2.14.12 Robustness comparisons for 1844 and 1846 122
Table 2.15.1 Sources of population and mortality data for 1841 123
Table 2.15.2 Population counts from the mustering of 1841 124
Table 2.15.3 Estimated population sizes for 1841: Ages 0-7 124
Table 2.15.4 Estimated population sizes for 1841: Ages 50+ 126
Table 2.15.5 Sampled and estimated number of deaths for 1841 127
Table 2.15.6 Estimates of current expected life for 1841 129
Table 2.15.7 Methodological comparisons for 1841 130
Table 2.15.8 Robustness comparisons for 1841 130
Table 2.16.1 Current expected life with standard error for the period 1841-1996 131
Table 2.16.2 Regression of the standard error of current expected life on population size 133

Table 3.5.1 Estimates of generation expected life for 1851, 1881 and 1901 144
Table 3.6.1 Components of generation lifetime distribution functions 1881-1900 146
Table 3.6.2 Average number of male civilian deaths per year 1915-19 147
Table 3.6.3 Number of war deaths of SA males from the Australian War Memorial 149
Table 3.6.4 Estimated number of war deaths of SA males 1915-19 151
Table 3.6.5 Estimated number of war deaths of SA males for generations 1881-1900 152
Table 3.6.6 Comparison of generation expected life of SA males 1881-1900 155
Table 3.6.7 Estimates of the number of embarkations of SA males 1915-18 157
Table 3.6.8 Proportion of embarkations of SA males for generations 1881-97 159
Table 3.6.9 Generation expected life of SA males with overseas military service 160

LIST OF FIGURES

Figure 1.5 Timescale 11

Figure 1.11.1 Triangular distribution T(a,m,b) 32
Figure 1.11.2 Comparison of distributions N(0,1) & T(-2.5,0,2.5).....34

Figure 2.16.1 Current expected life for the period 1841-1996.....134
Figure 2.16.2 The relationship between standard error and total population size.....135

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.