Technical Report
Calculating Life Expectancy in small areas
Technical Report

Calculating Life Expectancy in small areas

This project was commissioned and funded by SEPHO and a steering group was set up to guide development and implementation.

Acknowledgments and thanks
Dr Ted Williams (WSHA), Dr Farhang Tahzib (WSHA), Dr Anita Sims (SEPHO), Graham Evans (WSHA), Daniel Eayres (National Centre for Health Outcomes Development), Myer Glickman (Office of National Statistics), Keith Hardman (West Sussex County Council), Victoria Axell and Natasha De Meric.
Additional thanks to the National Portrait Gallery, Professor Chiang, David Pinder (Woking PCT) and David Osbourne (Croydon PCT).

Dr Edward S Williams

Hywell Dinsdale

Daniel Eayres

Dr Farhang Tahzib

Directorate of Public Health
Adur, Arun and Worthing tPCT

Published by the South East Public Health Observatory
Chapter 1: Introduction
1. Preface
2. Objectives
3. Work programme
4. Project outputs

Chapter 2: A historical review of life expectancy calculations
1. Summary
2. Early developments
3. International Life Expectancy figures
4. Technical advances
5. Life expectancy figures by regional health authority
6. Life expectancy by local authority and health authority
7. Life expectancy variation within health authority – electoral ward analysis
References

Chapter 3: Calculating Life Expectancy
1. Summary
2. The Current Life Table
3. Life table terms
4. Methodological considerations
5. Comparing methodologies
6. Results
7. Discussion
8. Conclusions
References

Chapter 4: Issues in Calculating Life Expectancy
1. Mortality data
2. Population data
3. Population life years at risk and standard error
4. LE by male and female vs LE by persons
5. Confidence intervals
6. Conclusions
References

Chapter 5: The Impact of Nursing Homes
1. Background
2. History and literature
3. Nursing home deaths
4. Impact of nursing home deaths on LE in West Sussex and Croydon
5. Nursing homes and LE in the South East
6. Nursing homes and residential care homes
7. Effects on SMRs
8. Means of adjusting for nursing homes
9. Conclusions
References

Appendix A: Calculating life expectancy
1. The current life table: the Chiang method
2. The Government Actuary’s Department (GAD) method
3. The Silcocks method
4. Problems arising from small numbers
5. Problems arising from death counts of zero
6. Study methodology
7. Results
8. Conclusions
9. Summary of recommendations
References

Appendix B: Counting nursing home deaths
1. The number of nursing homes in a ward
2. The number of nursing home residents or beds
3. Identifying individuals in nursing homes using Exeter and mortality data
4. Nursing home population and deaths of nursing home resident
5. Identifying Nursing home deaths using mortality files
References
Introduction

1. Preface

In December 2001, Dr Farhang Tahzib and Dr ES Williams from West Sussex Health Authority (WSHA) approached the South East Public Health Observatory (SEPHO) with a proposal to investigate the feasibility of calculating life expectancy in small areas such as electoral wards. After discussion SEPHO agreed to collaborate with a research project into life expectancy.

2. Objectives

The initial objectives of the project were as follows:

1. To investigate and quantify the statistical robustness of electoral ward level life expectancy figures;
2. To determine which methodology best suited small populations;
3. To investigate and quantify differences in alternative estimates of size of electoral ward populations and their effect on life expectancy;
4. To investigate and quantify the effect of nursing homes on electoral ward level life expectancies;
5. To calculate life expectancy figures for all wards in the South East region.

3. Work programme

The team investigated the detailed methodological issues surrounding life expectancy (LE) calculations, made contact with analysts around the South East and obtained extracts from the Exeter database for the majority of health authorities in the South East.

Methodological issues were then modelled using a spreadsheet-based life table template developed to test variations on the chosen methodology (e.g. using different age bands) and to test the robustness of this model in the face of variations in population, deaths or death rates. Data for all wards in West Sussex were used for this analysis.

Once the optimum methodology had been identified and its sensitivity understood, the differences between using general practice registration data (the Exeter system) and data from the 2001 Census were investigated.

The impact of nursing home deaths on life expectancy calculations was also examined. As part of the nursing home analysis, data for Surrey and Croydon were used in addition to data for West Sussex; local expertise in these places meant that nursing home data were available that could not be obtained for other areas.

In the final stages the project focused on calculating life expectancy figures for all wards in the South East Region, using the Census 2001 population and mortality data from 1998-2002. The work of the project has resulted in two papers being published in the Journal of Epidemiology and Community Health.

These methods have been adopted by the Association of Public Health Observatories (APHO) as their recommended approach for calculating life expectancy.
4. Project outputs

1. A seminar was held in London in March 2003 to discuss the preliminary findings of the project.

2. The proceedings and presentations from this event have been written up and made available on the SEPHO website.


4. A paper on ‘The impact of nursing home deaths on small area life expectancy calculations’ has been accepted for publication by the JECH.
   Williams ES, Dinsdale HN, Eayres D, Tahzib F. Impact of nursing home deaths on life expectancy calculations in small areas Journal of Epidemiology and Community Health 2004; 58: 958 - 962

5. A poster presentation was made at the 4th South East Public Health Conference in February 2004 at the Royal Society in London.

6. A write up of the project appeared in the Director of Public Health for AAW PCT’s Annual Report 2003

7. A set of LE figures for wards and Local Authorities across the South East region, calculated using deaths for 1998-02 and Census 2001 populations, have been made available through the SEPHO website.

8. An Excel spreadsheet based ‘calculator’, to facilitate small area life expectancy calculations using the method recommended by this project, has also been published through the SEPHO website.

9. Adoption of methods as the APHO recommended approach for calculating small area life expectancy.
A historical review of life expectancy calculations

1. Summary
Life expectancy was one of the earliest measures of mortality; deaths were first registered in England in 1603 and Edmund Halley produced the first life table by the end of the seventeenth century. In 1839 William Farr, the Registrar General of Births, Deaths and Marriages, published the first English National Life Table. Such life tables are still published today by the Government Actuary’s Department. Farr also used life expectancy figures for regional areas to emphasise health inequalities between different areas and occupational groups – highlighting geographical health inequalities.

Life expectancy is commonly used to make international comparisons, but has been used less with smaller populations. In the 1970s figures were published for regional health authorities and by the 1990s life expectancy figures were regularly being calculated for local authority areas. More recently there has been an interest in variation within local authorities, and the Office of National Statistics is considering making ward level figures routinely available.

2. Early developments
The story of life expectancy (LE) calculations probably starts in England in 1603, the first year of the reign of King James, when weekly bills of mortality were regularly collected for the first time. John Graunt (figure 1), a London merchant, took a lively interest in mortality and in 1662 published his Observations on the Bills of Mortality (figure 2).1 In his preface Graunt made the observation that the people who collected the weekly bills on mortality made little use of them. However Graunt’s curiosity was aroused and he examined the bills ‘so as to have a view of the whole together, in order to the more ready comparing of one year, season, parish, or other division of the city, with another, in respect of all burials, and christenings, and of all diseases and casualties happening in each of them respectively’. From these observations Graunt produced tables of mortality. He then tried to interpret the data in order to find some truths and to consider what benefit the knowledge would bring to the world. Graunt is widely recognised as one of the fathers of epidemiology and his work laid the foundation for more sophisticated analysis.

The famous English astronomer Edmund Halley (figure 3), who discovered Halley’s Comet, was also the inventor of the life expectancy table in 1693 (figure 4). His work allowed a probability based life expectancy to be derived from mortality tables. He wrote ‘An Estimate of the Degrees of the Mortality of Mankind, drawn from curious Tables of the Births and Funerals at the City of Breslaw; with an Attempt to ascertain the Price of Annuities upon Lives’. Halley pointed out that Graunt was not relating the number of deaths to the size of the population; in other words, he was not using a denominator. Halley wrote: ‘First, the number of the people was wanting. Secondly, the age of the people was not to be had.’ He was also aware of the bias caused by migration. Halley was able to calculate life tables for Breslaw, capital of the German province of Silesia, because the age and sex of death was collected, and the population was known.
Halley’s work was used by the English government to calculate LE tables and to sell annuities and marine insurance; it led to the founding of Lloyd’s of London. But Halley was more than a brilliant scientist. He thought deeply about the meaning of what he was doing and did not want his work to be misused. He wrote, ‘we think ourselves wronged if we attain not old age; whereas it appears thereby, that the one half of those that are born are dead in seventeen years time, 1,238 being in that time reduced to 616. So that instead of murmuring at what we call an untimely death, we ought with patience and unconcern to submit to that dissolution which is the necessary condition of our perishable materials, and of our nice and frail structure and composition. And to account it as a blessing that we have survived, perhaps by many years, that period of life, whereat the one half of the whole race of mankind does not arrive.’

The next person to make a significant contribution to the analysis of mortality data was Benjamin Gompertz (figure 5). Gompertz was a self-educated mathematician who learned his profession by reading Newton and acquired the skills to apply the calculus to actuarial questions. In 1825 he showed that age specific mortality rates increase in a geometrical progression, so when death rates are plotted on a logarithmic scale they result in a straight line, known as the Gompertz’s Law of Mortality. William Makeham (1867) modified the Gompertz function by adding a time independent parameter to represent the effect of chance environmental events, uncorrelated with age, that increased mortality risk. Makeham assumed that death was the consequence of two generally coexisting causes: chance, and a deterioration or increased inability to withstand destruction. The first, he argued, was a constant, while the second increased by geometrical progression.

In the nineteenth century two Victorians, Thomas Edmonds and William Farr, played a prominent role in the developing of vital statistics and the use of statistics to assess health. Edmonds was an actuary, who like Farr was concerned about the problems of urban poverty and the health of the industrial working class. Both believed that these problems could be studied objectively. Edmonds, as a professional statistician in the insurance industry, published many articles in *The Lancet* and wrote a book, *Life tables founded upon the discovery of a numerical law regulating the existence of every human being*. He demonstrated that human mortality varied in geometrical series in three periods of life that he labelled infancy, manhood and old age. He calculated for each year of life a mortality rate from which he could estimate the number remaining alive at each age. Unfortunately Edmonds failed to properly acknowledge the work of Gompertz and this led to a bitter controversy. Nevertheless, Edmonds’ importance resides in the applications he found for the law of mortality. Following the publication of the 1831 census, Edmonds produced age specific mortality rates as an indicator of general health. He constructed tables using the law of mortality for each county of England.
Meanwhile in 1839 William Farr (figure 6) had become the compiler of abstracts at the newly established General Register Office, a post he was to hold for the next forty years. He had at his fingertips an unprecedented quantity of data from the system of mortality registration that had been set up in 1837. Building on the work of Edmonds, Farr's task was to develop a system to analyse mortality data. He calculated age specific mortality rates and life expectancy, and produced the first national English Life Table.6

The first English Life Table was constructed in 1839 using only registered births and deaths since, in the opinion of William Farr, census figures at that time were unreliable.7 As the 1841 census provided more accurate data on the age and sex structure of the population, the second English Life Table was produced using mortality for the period 1838-44 and based on the 1841 census population. Farr constructed life tables for the entire population which appeared in the Registrar-General’s fifth and sixth annual reports. He also produced life tables for different areas and occupational groups. In the fifth report Farr showed that in Liverpool about half of all individuals had died by around age 5, in London by age 40, and in Surrey by age 55. The life expectancy for males in the three areas was 25 years for Liverpool, 35 years for London and 44 years for Surrey.8 In each area the life expectancy for women was around 2 years longer than for men. Farr pointed out how conditions in large urban areas adversely affected human health.

The Registrar’s Office then produced English Life Tables every few years until the end of the nineteenth century. From the first decade of the twentieth century life tables were produced on mortality data from the three years around a census year. The eighth Life Table was produced for the three-year period 1910-12, using population data from the 1911 census. In the 1920s the Government Actuary’s Department took over the preparation of the English Life Table and produced the ninth English Life Table based on the 1921 census.9 The Government Actuary’s Department also became responsible for providing life tables for Scotland and Northern Ireland. The fifteenth English Life Table was produced for the period 1990-92.

Towards the end of his career in 1875, Farr wrote ‘How the people of England live is one of the most important questions that can be considered; and how - of what causes, and at what ages - they die is scarcely of less account; for it is the complement of the primary question teaching men how to live a longer, healthier and happier life.’

The Government Actuary’s Department now produces interim life tables annually for the UK and its constituent countries. The interim tables are based on population estimates and mortality data for a period of three consecutive years.

3. International Life Expectancy figures

During the nineteenth century the use of life tables spread to other European countries, particularly Scandinavia. In the United States
official complete life tables were first produced in 1900-02 in connection with their decennial population census. Since 1939-41 these have been available for each of the states and since 1959-61 by state and race. In 1945 the US started a series of annual abridged life tables, based on annual death registrations and post-censal population estimates, that has continued to the present day.11

The proliferation of the use of life tables to other countries has been hampered by the availability of good quality vital events registration data and reliable population estimates. In 1968, Keyfitz and Flieger published a compilation of life tables for a large number of countries where the official data was of satisfactory quality.12 It covered only 29 per cent of the world’s population, predominantly in Europe and North America, with little representation from developing countries.

This lack of statistical data has prompted the development of indirect techniques for obtaining mortality rates, often based on observed similarities in the age patterns of mortality for different populations, such as those made by Gompertz. This involves using one or more parameters from the limited data that are available to select one of a number of empirical reference life tables which can then be modified to better suit the population in question. The United Nations published the first set of model life tables in 1995.13 The model was constructed using 158 reference life tables for each sex and statistical techniques to relate mortality at one age to mortality at another age over a range of mortality levels. Other series of model life tables were published by Coale and Demeny,14 Ledermann15 and Brass.16

In 1981 the United Nations published a set of life tables explicitly for use in developing countries17 and since 1999 the WHO has constructed annual life tables for all Member States using a modified version of the Brass model and taking particular account of the effect of HIV/AIDS on the pattern of mortality.18

4. Technical advances

In addition to these developments in model life tables, the twentieth century also saw continual refinement of the statistical techniques used in the construction of empirical life tables. Of primary concern is the fundamental step in life table construction whereby the observed age specific death rates are converted into their corresponding probabilities of dying. This is particularly important for abridged life tables where, generally, 5 or 10-year age groups are used. In 1939 Reed and Merrell proposed a formula for calculating the required probabilities based on a statistical analysis of the 1910 series of United States Life Tables.19 Greville took a different approach starting with Gompertz’s Law of Mortality to mathematically derive his formula.20 A relatively simple method was published by Weisler in 1954 whereby the observed death rates were substituted directly for Gompertz’s Law of Mortality to derive his formula.21 Keyfitz suggested a more complex approach involving an iterative process that repeats until the life table age specific rates agree with those observed. The
method suggested by Chiang (figure 7) in 1968\(^2\) is one of the most widely cited and is used in the UK by the Office for National Statistics for its sub-national life expectancy estimates.

Developments also took place in methods used to calculate the variance of the observed life expectancy. The variance is an essential quantity if statistical testing of redundant comparisons between the life expectancies of different populations are to be undertaken. Methods were published by Wilson in 1938\(^2\) and Irwin in 1949.\(^2\) In the 1960s Chiang investigated the distribution of the life expectancy estimate,\(^2\) and included a formula for the variance in his methods for life table construction.\(^2\) More recently, the WHO used Monte Carlo simulation techniques to estimate the 95% confidence limits of its model life table based life expectancies for all member states.\(^2\)

5. Life expectancy figures by regional health authority
In the 1970s, Gardner and Donnan used abridged life table techniques to compare LE among hospital regions within England and Wales.\(^3\) They argued that the use of life table techniques, not previously employed to compare health regions, allowed differences to be expressed in units of years, rather than using the dimensionless Standardised Mortality Ratios (SMRs). They systematically looked at life expectancy figures for each regional health authority in England and Wales from the years between 1974 and 1975. Results were presented as an alternative measure to SMR and illustrated substantial differences in the average years of expected life. The expectation of life ranged from 67.9 years for males and 74.3 years for females in the North Western regional health authority to 71.3 years for males and 76.9 years for females in the East Anglian regional health authorities. This analysis confirmed that people living in rural and prosperous areas tend to have the highest life expectancy whereas people living in urban inner city areas tend to have the lowest life expectancy. Gardner and Donnan described a north/south divide between English regions, with a regional variation of 3.4 years for males and 2.6 years for females.

6. Life expectancy by local authority and health authority
In 1996 Charlton argued that regional health authorities were large and each had a very heterogeneous composition in terms of population characteristics. It followed that examining data for such large areas may mask real differences within regions. Charlton analysed mortality data for the period 1981-92, producing life expectancy figures by local authority for the whole of England and Wales. The study showed significant variations in life expectancy within regional health authority areas. People living in rural and prosperous areas tend to have the lowest risk of mortality and those in urban, inner city areas the highest.\(^3\) Life expectancy figures for local authorities are now common and are currently the smallest units for which the Office for National Statistics routinely publish life expectancy\(^3\), using the method described by Chiang.\(^3\)
In 1997 Raleigh and Kiri stated that ‘life expectancy figures are so far available nationally, for regional health authorities, and for “clusters” of homogeneous areas but have not hitherto been examined for health authorities’. There is a need for such analysis in order to better understand intra-regional variation in health. Using mortality data for the decade 1984-1994 Raleigh and Kiri analysed trends in life expectancy across time for the 104 health authorities in England. Results showed that life expectancy in health authorities varied by 6.7 years in males and 4.7 years in females. By examining gender differences and trends in LE over the decade by health authority and level of deprivation, Raleigh and Kiri demonstrated that life expectancy in English district health authorities was inversely associated with deprivation. Moreover, the greatest gains over the decade were among the prosperous, longest-lived populations. Raleigh and Kiri stated that their findings were essential in enabling a better understanding of geographical health variation but also pointed out that health differentials were likely to be present even within health authorities.

In 2001 Silcocks et al used a model population of 256,000 to investigate the sampling distribution and usefulness of LE at health district level and below. This work confirmed that life expectancy was a valid measure for comparing mortality between populations of this size. In the same year Griffiths and Fitzpatrick examined geographic variation in LE in the UK for the ONS publication Health Statistics Quarterly. They produced abridged life tables for local authorities, but excluded from their analysis the Isle of Scilly (population 2,100) and the City of London (population 7,100) ‘as there are too few deaths there in a three-year period to make analysis meaningful’.

7. Life expectancy variation within health authority – electoral ward analysis

Acknowledging the wide internal variations in mortality that existed within large populations such as health districts, Townsend et al suggested an analysis of mortality at ward level, and showed variation in mortality rates between wards in northern England. Bremner put forward a similar argument a few years later, claiming there is ‘a need to move away from the convention of reporting (life expectancy) only for large, socially heterogeneous aggregates such as health or local authority areas...because such aggregates obscure differences between social areas’.

In 1995 Williams et al demonstrated this principle by calculating abridged population life tables for electoral wards within the London borough of Croydon to reveal differences of 5.4 years between the highest and lowest wards. Life expectancies for these 27 electoral wards were also presented in the annual public health report. They also showed that nursing home deaths were a potential source of bias and concluded that ‘it is therefore essential that an analysis of SMRs at electoral ward level takes account of nursing home deaths’.

Although the government has set an inequalities target in terms of life
expectancy (to reduce the gap between the quintile of areas with the lowest life expectancy at birth and the population as a whole), this is currently monitored at Health Authority level. The Department of Health has stated that life expectancy data at the ward level may not be sufficiently meaningful or robust, being strongly influenced by factors such as the number of old people’s homes in the ward and the problems of obtaining accurate and up to date ward populations.43

Recently, the former Trent Public Health Observatory (now the East Midlands Public Health Observatory) has produced LE calculations at ward level using Silcocks’ methodology44, and the 2000 West Sussex Public Health Report identified the ten electoral wards with the shortest and longest life expectancy.45 In late 2003 the ONS reported on methodological options for calculating LE in small populations.46 It is likely that ward figures will routinely be made available in the future.
References

2. Edmund Halley, Some further considerations on the Breslaw Bills of Mortality, cited from And Estimate of the Degrees of the Mortality of Mankind, drawn from curious tables of births and funerals at the City of Breslaw, with an attempt to ascertain the price of annuities upon lives. By Mr Edmund Halley, edited by Matthias Bohne, January 2002, Halley’s postscript, p654-56
5. John M Eyler, Constructing vital statistics; Thomas Rowe Edmonds and William Farr, 1835-1845, Department of History of Medicine, University of Minnesota, Series: History of Epidemiology,
10. William Farr, letter to the Registrar General on the mortality in the registration districts of England during the years 1861–70, cited from supplement to the 35th annual report of the Registrar General of births, deaths and marriages in England, HMSO, 1875
27. Ibid. Chiang CL. The life table and its construction. II. Human Biology, 1960;32:221-238.


38. Townsend, P Phillimore P and Beattie, A. *Health and Deprivation: Inequality and the North* 1987 London Croom Helm


42. Ibid. Using mortality data to describe geographic variations in health status at sub-district level. *Public Health* 1995; 109: 67-73


45. West Sussex Public Health Report, Dr S Parker, et al, ISBN.

Calculating Life Expectancy

This chapter provides a summary of the methodological issues investigated by the Small Area Life Expectancy Project. Much of the work is based on a model, designed in Excel by Daniel Eayres, which enabled multiple Life Expectancy (LE) calculations to be performed automatically for a number of methodological variations. The model shown to be most appropriate for small area LE calculations is detailed in this chapter.

A more detailed account of life table theory and both the methodology and the results of this project are available in appendix A. The outputs of the project’s methodology work have been peer reviewed and published in the Journal of Epidemiology and Community Health.

1. Summary

- The current life table is the most effective way of summarising the mortality experience of a population and can also be used to make statistical inferences and comparisons between the mortality experiences of different populations.
- Life tables may be complete or abridged. For small areas practicality and the problem of small numbers rule out complete life tables so abridged life tables must be used. The common terms used in abridged tables are described in this chapter.
- Using Monte Carlo simulations for populations of various sizes, the project compared life expectancy figures produced using the two most common life table methodologies of Chiang and Silcocks.
- The effect of different age intervals, the calculation of confidence intervals and the impact of zero deaths within an age band was also investigated.
- The Chiang methodology was shown to produce better estimates of LE for small populations. It was also shown that models with 5-year age bands to 85+ performed best.
- The life expectancy estimates produced showed a normal distribution even for small populations. The normal approximation for confidence intervals is valid.
- Zero deaths were shown to have a negligible effect on the standard error of the LE estimate. However, a zero death count in the final end band leads to an infinite life expectancy so the appropriate national mortality rate must be used in this instance.
- Standard error (and hence confidence limits) increase as population decreases. A population of 5,000 life years at risk produces an ‘acceptable’ standard error of +/- 2 years (or a 95% confidence limit of +/- 4 years).
2. The Current Life Table

The calculation of life expectancy requires the construction of a table of information that breaks down the mortality experience of a population by age. This kind of table is referred to as a life table and there are two principal types: the cohort life table and the current life table.

The cohort life table records the actual mortality experience of a cohort of individuals from the birth (or other event such as diagnosis of cancer) of the first member to the death of the last. Such tables are susceptible to loss to follow up and only produce life expectancies for a cohort that is already dead. Cohort life tables are useful for studying patient survival after treatment or intervention.

The current life table gives a cross sectional view of the mortality experience of a population during a given time period. It depends solely on the age-specific mortality rates prevailing in the population during the time period under consideration. These rates are applied to a hypothetical cohort of newborns to calculate their average expected life span or life expectancy. Life expectancy can therefore be defined as the average number of years a newborn can expect to live if it experiences the currently prevailing age-specific mortality rates throughout its life span. Life expectancy at other ages can be calculated in a similar way, for example, life expectancy at age 65 can be calculated by applying the current age specific mortality rates over the age of 65 to a hypothetical cohort of 65 year olds. The current life table is the most effective way of summarising the mortality experience of a population and can also be used to make statistical inferences and comparisons between the mortality experience of different populations. The life expectancies calculated in this study are from current life tables.

Both cohort and current life tables may be complete (that is, unabridged) or abridged. In complete life tables the mortality experience is broken down by individual years of life and occasionally even further, particularly for mortality under 1 year of age where most infant deaths occur. In abridged life tables the mortality experience is broken down by larger age intervals, usually of 5 or 10 years, with the general exceptional of the first year of life, where it is usual to have a separate age interval due to the relatively high death rate that occurs in the first year of life. For small areas, practicality and the problem of small numbers rule out complete life tables so abridged life tables must be used.

3. Life table terms

Table 1 shows an example of an abridged current life table, using the mortality data for males in England over the period 1998-00. The table has been completed using the methods described by Chiang. This methodology is used by The Office for National Statistics to produce life expectancies at Health Authority and Local Authority level. It has also been used by the London Public Health Observatory in their study of the Health Inequality Targets as applied to London.
A brief description of each of the terms used in the life table is given below. A more detailed description is given in appendix A.

- **Age interval**
The period of life between two ages stated in years. For example, 10-14 is the 5-year age interval between the 10th and 15th birthdays.

- **Interval width**
The width in years of the age interval.

- **Fraction of last age interval of life**
Members of the hypothetical cohort who die during an age interval do not all do so at either the beginning or the end of the interval, but at various points through its length. The fraction of last age interval of life is the average fraction of that age interval that they survive before dying. This term is not shown in table 1.

- **Population years at risk**
The population years at risk for the age interval in the study population. For example, for the 10-14 age interval in table 1, the sum of the mid-year population estimates of English males aged 10-14 years for the years 1998, 1999 and 2000 has been used, as three years' mortality data was analysed to obtain the number of deaths.
CHAPTER 3

Life Expectancy Report 2005

- Number of deaths in interval
  The number of deaths observed in the age interval in the study population.

- Annual death rate in interval
  The average annual mortality rate for that age interval in the study population, based on the observed number of deaths. This is found by dividing the number of deaths in interval by the population years at risk.

- Probability of dying in interval
  The proportion of the hypothetical cohort alive at the beginning of the age interval who will die during the age interval. For example, in Table 1 the probability of dying in the first year of life is 0.00628, that is for every 100,000 newborn babies 628 will die before their 1st birthday. This probability of death can be derived from the observed mortality rate (the annual death rate in interval), the interval width and the fraction of last age interval of life. Since the final age interval is open-ended, the probability of dying in interval is by definition equal to 1.

- Number alive at start of interval
  The number of the hypothetical cohort alive at the start of the age interval. The size of the cohort at birth is arbitrary and is usually set to 100,000. The number alive at the start of subsequent age intervals can be calculated by subtracting the number dying in interval from the number alive at start of interval for the previous age interval.

- Number dying in interval
  The number of the hypothetical cohort dying in the age interval (that is, the expected number of deaths). It can be calculated by applying the probability of dying in interval to the number alive at start of interval. As the final age interval is open-ended all those alive at the start of the interval will die during it.

- Number of years lived in interval
  The number of person years lived during the age interval by the members of the cohort who are alive at the start of the interval. Those who survive the age interval each contribute to the whole interval width. Those who die during the interval each contribute on average a proportion of the interval width, determined by the fraction of last age interval of life.

- Total number of years lived beyond start of interval
  The total number of person years still to be lived by members of the cohort who are alive at the start of the age interval. It is the sum of the number of years lived in interval for the current age interval and all the subsequent intervals.

- Observed expectation of life at start of interval
  The average number of years that each of the members of the cohort alive at the start of the interval can expect to live. It is calculated by dividing the total number of years lived beyond start of interval by the number alive at start of interval. This is the Observed Expectation of Life or Life Expectancy. For example, in Table 1 a newborn baby can expect to live for 75.4 years and one who reaches his 65th birthday can expect to live another 15.6 years.
Further columns are required if the variance or 95% confidence interval of the life expectancy estimate, or other statistical comparisons are to be calculated. Details can be found in appendix A.

4. Methodological considerations

There are a number of different published methodologies for completing a life table, many of which are referenced and discussed in the previous chapter. The principal difference between these methods lies in how they convert the observed age-specific mortality rates into the age-specific probabilities of dying applied to the hypothetical cohort. The exact nature of the relationship depends on the assumptions that the particular methodology makes.

Two main methods were considered for this project: the Chiang method (as previously described) and the Silcocks method. A full description of the two methods is given in appendix A. The impact on both of using different age bands or different endpoints and the calculation of the standard error for estimates produced with the two methods was investigated. The impact of zero deaths on the estimation of standard error, a particular problem when conducting studies on small areas, was also assessed.

5. Comparing methodologies

5.1 Comparing the Chiang and Silcocks methodologies

Both the Chiang and Silcocks LE methodologies are based on the construction of a current life table. While Chiang uses a linear method that assumes deaths are distributed evenly through an age interval, Silcocks assumes that the mortality rate is constant throughout an age interval, resulting in the number of survivors decreasing exponentially (see figure 1).

5.2 Investigating the effect of different age intervals

Does the use of different age intervals in the life table calculations have a significant effect on the resulting life expectancy estimates? The effect of different age intervals on the estimates of LE was investigated, examining 5-year versus 10-year age intervals and different final-age intervals (85+, 90+, 95+).
5.3 Calculation of confidence intervals

At electoral ward level the populations and number of deaths involved are relatively small: the average English electoral ward has a population of around 6,000 persons and experiences about 60 deaths a year. The standard errors of life expectancies calculated from such numbers are likely to be relatively large and the confidence intervals wide.

In England, confidence intervals have not commonly been quoted for LE estimates, even for sub-national areas such as Local Authorities. The ONS published LE estimates with confidence intervals for the first time in 2003. Both Chiang and Silcocks provide a formula for calculating the standard error but their assumptions differ. Chiang assumes that observed age-specific deaths are binomially distributed, while Silcocks assumes a Poisson distribution. In age intervals where death is a rare event the two distributions are approximately equivalent. However, for some age intervals the magnitude of the mortality rate is such that death cannot be considered as a rare event and there is some debate as to which distribution is the most appropriate to use. Each methodology was tested using deaths simulated from the distribution upon which that methodology is based.

A further difference between the two methods is the variance term for the final-age interval. Chiang assumes that since the probability of survival in this interval is by definition zero, the associated variance is also zero. Silcocks et al argue that for the final-age interval the LE is dependent, not on the probability of survival, but on the mean length of survival, and have included a term for the variance based upon this assumption. The Silcocks argument is thought to be valid and so the additional variance term has been included within the Chiang methodology during some of these investigations. This is reflected in the results by the label Chiang (Adjusted). Note that the adjustment affects the estimate of the standard error only and not the LE estimate itself.

5.4 Investigating the effect of zero death counts

Another problem associated with small populations is the occurrence of zero deaths in an age interval which is particularly important in the final-age interval. If there are no observed deaths the death rate ($M/\ell_{x}$) is zero and the hypothetical cohort surviving to the start of the final-age interval will have an infinite mean length of survival ($1/M/\ell_{x}$) and an infinite LE. In such instances an alternative rate must be used, such as the appropriate national or regional rate or the (weighted) average rate of the surrounding areas. In this study, the equivalent death rate for England has been used in all models tested.

Silcocks noted that a zero count gives an estimate of zero for the sample variance of the age interval, which is an underestimate of the true variation. This results in an underestimation in the total life table variance and therefore of the LE standard error. The greater the number of zero death counts, the greater the underestimation of the standard error. Silcocks suggested two possible values, 0.693 and 3.0,
as possible substitutes for zero counts. These values are the Poisson means for which the probability of observing zero deaths is 50% and 5% respectively. A third option was also included in the simulation: the expected number of deaths based on the appropriate national or regional age-specific mortality rate. In this study, the equivalent age specific death rate for England has been used.

5.5 Study Methodology
Monte Carlo simulations were performed for a hypothetical population of varying size to describe the distribution of the LE estimate and its standard error estimate. The population age structure and the underlying age-specific mortality rates of the population were set to those of English males 1998-00, allowing comparisons to reference LE figures published by the Government Actuary’s Department.12

A single simulation consisted of 10,000 repetitions. For each repetition, the underlying age-specific mortality rate and the population for each of the age intervals of the life table model being tested were used to generate a random count of deaths from a known probability distribution. These counts were inserted into a life table to generate an estimate of LE and its standard error. The results of the repetitions produced distributions of LE and standard error estimates from which inferences could be made.

Simulations were performed for each methodology, for 5 and 10-year age intervals, to the final-age intervals of 85+.90+.95+, or 85+.95+, respectively. The first years of life were broken down into the age groups under 1 and 1-4 years. Simulations using Chiang assumed the Binomial distribution for generating random counts of death, while those using Silcocks assumed the Poisson distribution. Simulations were repeated for the hypothetical population-years at risk: 500; 1,000; 5,000; 10,000; 25,000 and 50,000.

To investigate the potential problem caused by zero deaths in an age interval, a count of such occurrences was recorded for each repetition. Simulations were also repeated for each of the following substitutions for zero deaths:
- none;
- values 0.693, 3 or the expected number of deaths in both the LE and the standard error calculation;
- values 0.693, 3 or the expected number of deaths in the standard error calculation only.

A Reference LE and standard error for each abridged life table model were calculated using the underlying mortality rates directly. An additional Unabridged Reference LE was calculated for each methodology using the underlying mortality rates in an unabridged life table. A Microsoft Excel application was developed to perform the simulations and the results were exported to SPSS for statistical analysis.

6. Results
6.1 Reference life expectancy
Table 2 shows the Reference LE for both the abridged model and the unabridged life table using Chiang and Silcocks methodologies. In addition the Government Actuary’s Department (GAD) figures are presented for comparison.
The Reference LE is independent of the population size as it is calculated from the exact underlying rates.

The abridged life tables using Chiang produced slightly better approximations to the GAD reference figure than those using Silcocks. For all the alternative age interval models, the Silcocks methodology produced higher expectancies than the Chiang methodology. For both methodologies 5-year age intervals resulted in better estimates than 10-year intervals.

### 6.2 Simulated life expectancy

Table 3 shows the mean of simulated life expectancies for each life table model for population-years-at-risk ranging from 500 to 50,000. Each mean is the average of the 10,000 life expectancies generated by the repetitions for that particular model and population size.

Logic would suggest that the mean simulated life expectancies would be independent of the population size but this is not the case (figure 2). For all life table models, smaller population sizes gave higher

---

**Table 2**

**Calculated Life Expectancy By Methodology Assuming Male Mortality Rates For England 1998-00**

<table>
<thead>
<tr>
<th>Method</th>
<th>Life Table</th>
<th>Life Expectancy</th>
<th>Difference From GAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government Actuaries Dept</td>
<td>Unabridged</td>
<td>75.42</td>
<td>0.00</td>
</tr>
<tr>
<td>Chiang</td>
<td>Unabridged</td>
<td>75.42</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Abridged 5Yr 85+</td>
<td>75.44</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>5Yr 90+</td>
<td>75.41</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>5Yr 95+</td>
<td>75.40</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>10Yr 85+</td>
<td>75.38</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>10Yr 95+</td>
<td>75.39</td>
<td>-0.03</td>
</tr>
<tr>
<td>Silcocks</td>
<td>Unabridged</td>
<td>75.42</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Abridged 5Yr 85+</td>
<td>75.49</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>5Yr 90+</td>
<td>75.50</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>5Yr 95+</td>
<td>75.51</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>10Yr 85+</td>
<td>75.53</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>10Yr 95+</td>
<td>75.79</td>
<td>0.37</td>
</tr>
</tbody>
</table>

---

estimates of the LE. As the population size increases, the mean simulated life expectancies approach their corresponding Reference LE. For populations greater than 5,000, the choice of methodology is the most important factor in estimating LE, with all Chiang models giving better approximations than any of the Silcocks models. In populations of 500, the most important factor is choice of the final-age interval, with models using the 85+ endpoint giving the best approximations to the GAD reference figure. This suggests that a significant proportion, if not all, of the ‘drift’ in the mean simulated LE for smaller populations might be found in the final-age interval.

6.3 Final-age interval
Both Chiang and Silcocks estimate the mean survival within the final-age interval by \(1/M_{85+}\), where \(M_{85+}\) is the mortality rate of the final-age interval. The left-hand side of figure 3 shows the distribution of 10,000 simulated \(M_{85+}\) for the 85+ age interval of a Chiang model, for population sizes of 50,000, 10,000 and 5,000. The distribution of the simulated rates is a weighted binomial (being the binomially distributed death counts divided by the 85+ population). As the population size increases, the distribution tends towards the Normal and the standard deviation decreases. The mean of the distribution, which is an estimate of the underlying mortality rate, is independent of the population size.

The right-hand side of figure 3 shows what happens when these simulated rates \(M_{85+}\) are transformed to give the simulated mean survival of the final-age interval, \(1/M_{85+}\). The right hand tail becomes stretched and, as the population size decreases, the skewing of the transformed survival time distribution increases, shifting the distribution mean increasingly

### Table 3
Mean Of Simulated Life Expectancies By Methodology And Population Size Assuming Male Mortality Rates For England 1998-00

<table>
<thead>
<tr>
<th>Method</th>
<th>Life Table</th>
<th>Reference Life Expectancy</th>
<th>500</th>
<th>1,000</th>
<th>5,000</th>
<th>10,000</th>
<th>25,000</th>
<th>50,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government Actuaries Dept</td>
<td>Unabridged</td>
<td>75.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiang</td>
<td>Unabridged</td>
<td>75.42</td>
<td>76.36</td>
<td>76.18</td>
<td>75.70</td>
<td>75.56</td>
<td>75.49</td>
<td>75.47</td>
</tr>
<tr>
<td></td>
<td>Abridged 5Yr 85+</td>
<td>75.41</td>
<td>76.87</td>
<td>76.01</td>
<td>75.63</td>
<td>75.55</td>
<td>75.45</td>
<td>75.44</td>
</tr>
<tr>
<td></td>
<td>5Yr 90+</td>
<td>75.40</td>
<td>77.27</td>
<td>76.27</td>
<td>75.54</td>
<td>75.47</td>
<td>75.45</td>
<td>75.43</td>
</tr>
<tr>
<td></td>
<td>5Yr 95+</td>
<td>75.38</td>
<td>76.26</td>
<td>76.11</td>
<td>75.64</td>
<td>75.50</td>
<td>75.42</td>
<td>75.40</td>
</tr>
<tr>
<td></td>
<td>10Yr 85+</td>
<td>75.39</td>
<td>77.43</td>
<td>76.44</td>
<td>75.60</td>
<td>75.49</td>
<td>75.45</td>
<td>75.41</td>
</tr>
<tr>
<td></td>
<td>10Yr 95+</td>
<td>75.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silcocks</td>
<td>Unabridged</td>
<td>75.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abridged 5Yr 85+</td>
<td>75.49</td>
<td>76.57</td>
<td>76.42</td>
<td>75.84</td>
<td>75.65</td>
<td>75.55</td>
<td>75.52</td>
</tr>
<tr>
<td></td>
<td>5Yr 90+</td>
<td>75.50</td>
<td>77.01</td>
<td>76.18</td>
<td>75.80</td>
<td>75.67</td>
<td>75.56</td>
<td>75.52</td>
</tr>
<tr>
<td></td>
<td>5Yr 95+</td>
<td>75.51</td>
<td>77.70</td>
<td>76.65</td>
<td>75.69</td>
<td>75.62</td>
<td>75.59</td>
<td>75.56</td>
</tr>
<tr>
<td></td>
<td>10Yr 85+</td>
<td>75.53</td>
<td>76.51</td>
<td>76.38</td>
<td>75.85</td>
<td>75.68</td>
<td>75.59</td>
<td>75.56</td>
</tr>
<tr>
<td></td>
<td>10Yr 95+</td>
<td>75.79</td>
<td>77.90</td>
<td>76.84</td>
<td>75.96</td>
<td>75.87</td>
<td>75.87</td>
<td>75.84</td>
</tr>
</tbody>
</table>

Source:
Interim Life Tables 1998-00, Government Actuaries Dept
ONS Annual District Mortality Extracts 1998 - 2000
Figure 3
Distribution of simulated mortality rate and mean survival time in the 85+ age interval by population size; assuming the population age structure and mortality rates of English males 1998-00

Mortality Rate (M)
Expected = 0.185 deaths / person-years at risk

Population: 50,000
Mean = 0.185
Std Dev. = 0.017
N = 10,000

Population: 10,000
Mean = 0.186
Std Dev. = 0.038
N = 10,000

Population: 5,000
Mean = 0.186
Std Dev. = 0.053
N = 10,000

Mean Survival Time (1/M)
Expected = (1/0.185) = 5.39 Years

Population: 50,000
Mean = 5.44
Std Dev. = 0.50
N = 10,000

Population: 10,000
Mean = 5.64
Std Dev. = 1.29
N = 10,000

Population: 5,000
Mean = 5.95
Std Dev. = 2.27
N = 10,000

Mortality Rate (M) = 0.185 deaths / person-years at risk

Mean Survival Time (1/M) = (1/0.185) = 5.39 Years

Simulated Mean Survival (Years)
further to the right. This results in an overestimate of the underlying survival time and consequently the LE.

Using 90+ or 95+ as the final-age interval, as opposed to 85+, generally results in larger drifts in the LE estimate for small populations. This is because the populations in these age intervals are smaller, resulting in greater skewing of the transformed survival times.

Similar effects can be expected in the other, finite, age intervals where the years of life lived during the interval is related to the probability of dying (Chiang) or probability of survival (Silcocks), both of which are transformations of the mortality rate \( M_i \).

### 6.4 Distribution of simulated life expectancy

Silcocks demonstrated that LE estimates are normally distributed. The results of this project confirm this finding, demonstrating that it remains true even for population sizes of 5,000 (figure 4). This is an important observation as the Normal approximation method used to calculate 95% confidence intervals for the LE (i.e. \( 95\%CI = \pm 1.96 \times \text{std err} \)) remains valid for small populations.

### 6.5 Life expectancy standard error

Table 4 shows the standard error of LE estimates by population size. Three measures of the standard error are given: the first is the ‘reference’, calculated using the exact underlying mortality rates; the second is the ‘observed’, found by measuring the standard deviation of the simulated distribution of LE estimates (for example, figure 4); and the third is the ‘mean estimated’, that is, the mean of the simulated distribution of standard error estimates. The latter is of particular interest because we are interested in the ‘mean estimated’ standard error. By describing the distribution of these simulated standard errors inferences may be made about how the standard error estimate of real data will behave.

Table 4 shows that the standard error increases as the population size decreases. For the given age structure and underlying mortality rates, the standard error increases from approximately 0.6 years for 50,000 population, to 1.4 years for 10,000 population and 4.3 years for 1,000 population. The Chiang (Adjusted) and Silcocks methodologies give similar estimates of the standard error. For populations over 1,000 the width of the age intervals and the choice of the final-age interval have little effect on standard error calculations.

Examination of the distributions of standard error estimates showed that for population sizes down to 10,000 the distribution closely follows the Normal. For smaller populations, however, the distribution becomes increasingly skewed (figure 5). The width of the distribution indicates the precision of the standard error estimate. For the population size of 5,000 the mean estimate of standard error is 1.96 years, but its associated standard deviation is relatively large at 0.40 years.

### 6.6 Standard error and zero deaths

A further feature of table 4 is the good agreement of the ‘mean estimated’ standard error to the ‘reference’ and ‘observed’ standard errors. As the population size decreases, zero deaths counts became more frequent. Whilst the ‘reference’
Figure 4
Distribution Of Simulated Life Expectancy By Population Size Using Chiang Methodology - 5 Year Abridged Life Table To 85+ Assuming Male Mortality Rates For England 1998-00

Table 4
Reference And Simulated Life Expectancy Standard Error By Methodology And Population Size
Assuming Male Mortality Rates For England 1998-00

<table>
<thead>
<tr>
<th>Methodology Model</th>
<th>Age Intervals</th>
<th>Source Of Estimate Of Standard Error</th>
<th>Life Expectancy Standard Error For Hypothetical Ward With A Population Of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>500</td>
</tr>
<tr>
<td>Chiang Abridged</td>
<td>5Yr 85+</td>
<td>Reference Standard Error</td>
<td>5.993</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>5.739</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>4.850</td>
</tr>
<tr>
<td></td>
<td>5Yr 90+</td>
<td>Reference Standard Error</td>
<td>6.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>4.978</td>
</tr>
<tr>
<td></td>
<td>5Yr 95+</td>
<td>Reference Standard Error</td>
<td>6.039</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.393</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.405</td>
</tr>
<tr>
<td></td>
<td>10Yr 85+</td>
<td>Reference Standard Error</td>
<td>5.996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>5.797</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>4.809</td>
</tr>
<tr>
<td></td>
<td>10Yr 95+</td>
<td>Reference Standard Error</td>
<td>6.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.306</td>
</tr>
<tr>
<td>Chiang Abridged</td>
<td>5Yr 85+</td>
<td>Reference Standard Error</td>
<td>6.120</td>
</tr>
<tr>
<td>(Adjusted)</td>
<td></td>
<td>Observed Standard Error</td>
<td>5.739</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.161</td>
</tr>
<tr>
<td></td>
<td>5Yr 90+</td>
<td>Reference Standard Error</td>
<td>6.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.182</td>
</tr>
<tr>
<td></td>
<td>5Yr 95+</td>
<td>Reference Standard Error</td>
<td>6.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.393</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.799</td>
</tr>
<tr>
<td></td>
<td>10Yr 85+</td>
<td>Reference Standard Error</td>
<td>6.132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>5.797</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.124</td>
</tr>
<tr>
<td></td>
<td>10Yr 95+</td>
<td>Reference Standard Error</td>
<td>6.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.550</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.707</td>
</tr>
<tr>
<td>Silcocks Abridged</td>
<td>5Yr 85+</td>
<td>Reference Standard Error</td>
<td>6.152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>5.824</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.190</td>
</tr>
<tr>
<td></td>
<td>5Yr 90+</td>
<td>Reference Standard Error</td>
<td>6.095</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.053</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.350</td>
</tr>
<tr>
<td></td>
<td>5Yr 95+</td>
<td>Reference Standard Error</td>
<td>6.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.489</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>6.012</td>
</tr>
<tr>
<td></td>
<td>10Yr 85+</td>
<td>Reference Standard Error</td>
<td>6.182</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>5.778</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.197</td>
</tr>
<tr>
<td></td>
<td>10Yr 95+</td>
<td>Reference Standard Error</td>
<td>6.175</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed Standard Error</td>
<td>6.434</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>6.007</td>
</tr>
</tbody>
</table>

1 Reference Standard Error - the standard error calculated by the life table using the exact underlying mortality rates
2 Observed Standard Error - the standard deviation of the life expectancies calculated by each iteration of the simulated life table
3 Mean Estimated Standard Error - the mean of the standard errors calculated by each iteration of the simulated life table.

Source:
Interim Life Tables 1998-00, Government Actuaries Dept
ONS Annual District Mortality Extracts 1998 - 2000
CHAPTER 3

Life Expectancy Report 2005

and ‘observed’ figures are unaffected by the problem of zero death counts, it was expected that the ‘mean estimated’ figure would increasingly underestimate the true standard error. This occurs to a small degree for a population size of 500, but it was encouraging that this underestimation was not evident at larger population sizes.

The ‘mean estimated’ standard error by the frequency of the zero death counts was examined and it was shown that repetitions with a higher frequency of zero deaths give lower estimates of LE standard error (figure 6). The effect holds for both methodologies and for various age structures, and remains evident, although to a lesser degree, at a population size of 50,000. The most plausible reason that this effect does not manifest itself in the overall ‘mean estimate’ is that a mechanism, similar to that which causes the overestimation of the LE itself, causes an overestimation of the standard error, counterbalancing the underestimation caused by the occurrence of zero deaths.

6.7 Substitution for zero deaths

As shown in table 5, none of the substitution methods tested (0.693, 3.0 and the expected number of deaths) produced better estimates, A. This is unsurprising given the unexpected robustness in standard error estimates.

---

**Figure 5**
Distribution of simulated life expectancy standard error by population size; using the Chiang methodology with a 5 year abridged life table to 85+ and assuming the population age structure and mortality rates of English males 1998-00

**Figure 6**
Simulated life expectancy standard error versus the occurrence of zero death counts in the simulated life table, by population size; using the Chiang methodology with a 5 year abridged life table to 85+ and assuming the population age structure and mortality rates of English males 1998-00
CHAPTER 3
Life Expectancy Report 2005

7. Discussion
This is the first study, as far as we are aware, to systematically evaluate the application of standard life expectancy (LE) methodologies to populations of small areas.

LEs estimated using the Chiang and the Silcocks methodologies showed good agreement. However, for use in calculating small area LEs in England the Chiang methodology is recommended. It gave the better estimates when compared to the GAD reference and is consistent with the methodology used by ONS for larger populations. For estimating the LE standard error, we propose that the Chiang methodology be adjusted to include a term for the variance associated with the final-age band, as suggested by Silcocks.

The choice of the final-age interval is important because LE becomes increasingly overestimated as the population size diminishes. This effect is greatest with a final-age interval of 95+, and least with a final-age interval of 85+. We propose that for small populations 85+ is used as

Table 5
Effect On Simulated Life Expectancy And Standard Error Of Substituting For Zero Deaths By Substitution Method And Population Size Using Adjusted Chiang Methodology - 5 Year Abridged Life Table To 85+
Assuming Male Mortality Rates For England 1998-00

<table>
<thead>
<tr>
<th>Substitution Methodology For Handling Zero Death Counts</th>
<th>Life Expectancy Statistic</th>
<th>Standard Error Of Life Expectancy For Hypothetical Ward With A Population Of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference Life Expectancy</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>Reference Standard Error</td>
<td>75.44</td>
</tr>
<tr>
<td>No Substitutions</td>
<td>Mean of Simulated LEs</td>
<td>76.36</td>
</tr>
<tr>
<td></td>
<td>Observed Standard Error</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>5.16</td>
</tr>
<tr>
<td>Life Expectancy and Variance - 0.693 Deaths</td>
<td>Mean of Simulated LEs</td>
<td>34.78</td>
</tr>
<tr>
<td></td>
<td>Observed Standard Error</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>9.26</td>
</tr>
<tr>
<td>Life Expectancy and Variance - 3 Deaths</td>
<td>Mean of Simulated LEs</td>
<td>7.56</td>
</tr>
<tr>
<td></td>
<td>Observed Standard Error</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>2.69</td>
</tr>
<tr>
<td>Life Expectancy and Variance - Expected Deaths</td>
<td>Mean of Simulated LEs</td>
<td>69.96</td>
</tr>
<tr>
<td></td>
<td>Observed Standard Error</td>
<td>3.95</td>
</tr>
<tr>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>6.12</td>
</tr>
<tr>
<td>Variance Only - 0.693 Deaths</td>
<td>Mean of Simulated LEs</td>
<td>76.36</td>
</tr>
<tr>
<td></td>
<td>Observed Standard Error</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>21.93</td>
</tr>
<tr>
<td>Variance Only - 3 Deaths</td>
<td>Mean of Simulated LEs</td>
<td>76.36</td>
</tr>
<tr>
<td></td>
<td>Observed Standard Error</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>31.65</td>
</tr>
<tr>
<td>Variance Only - Expected Deaths</td>
<td>Mean of Simulated LEs</td>
<td>76.36</td>
</tr>
<tr>
<td></td>
<td>Observed Standard Error</td>
<td>5.74</td>
</tr>
<tr>
<td></td>
<td>Mean Estimated Standard Error</td>
<td>7.71</td>
</tr>
</tbody>
</table>

*Expected deaths is the number of deaths expected in the age interval if it experiences the exact national age-specific mortality rate
the final-age interval. For other age bands the choice between 5 or 10-years age intervals is more arbitrary and depends upon the availability of accurate denominator information. We recommend that, within England, 5-year age bands are used to retain consistency with ONS methods.

The increasing number of zero death counts within the life tables of small populations did not affect estimates of LE standard error to the degree anticipated. Methods of correcting standard error estimates by substituting zero death counts gave no better results than the simple uncorrected method. We recommend therefore that no substitutions are made for zero death counts, except where this occurs in the final-age interval. Such a count, if not substituted, would lead to an infinite LE. In such instances an appropriate national, regional, or locally derived age-specific mortality rate should be used.

Estimates of LE are normally distributed even for very small populations. As population size decreases the standard error increases and it becomes increasingly difficult to show statistically significant differences between areas. A population-years at risk of 5,000, with the same age structure and mortality rates as England males in 1998-00, has a LE standard error of approximately 2 years, giving a 95% confidence interval of ±4 years. For a population of 1,000 this interval rises to over ±8 years. To put this into context, there is a difference of approximately 8.5 years between the highest and lowest English Local Authority male life expectancies in the period 2001-03. For smaller populations this differential will be greater but it is clear that for English populations smaller than around 5,000 only those at the extremes of the range will show statistical significance.

The estimate of the standard error is subject to sampling variation that increases as the population size falls. The 95% confidence interval of the estimated LE standard error, quoted above for a population of 5,000, is itself a relatively large ±0.8 years. A further problem with estimating standard error of small populations is that its distribution becomes increasingly skewed.

8. Conclusions
When applying standard life expectancy methodologies to increasingly small populations the problems of overestimation of the LE, increasingly wide confidence intervals and increasingly poor estimation of its standard error must be carefully considered. For small areas in England, we suggest that a population-years at risk of 5,000 is a reasonable point above which LE calculations can be performed with confidence. For many areas it will be necessary to aggregate data either geographically or over time, particularly if sex-specific life expectancies are required.
CHAPTER 3

References

1 Eayres D and Williams ES Evaluation of methodologies for small area life expectancy estimation J Epidemiol Community Health 2004; 58: 243-249
Life expectancy (LE) calculations require only two forms of data: mortality and population. For small areas in the UK it is the latter measure that tends to cause difficulty as ward level populations are currently only available for the Census year. One alternative source of population data available to those in the health sector is provided by the Exeter database which records patients registered with GPs within an area. This chapter discusses the mortality and population data required for LE calculations and presents the findings of the analysis of Exeter data for the South East in terms of LE calculations.

In order to minimise the standard error of (and thus the confidence limits around) LE estimates, an adequate effective population size must also be achieved for each area (5,000 population years at risk is recommended). In addition decisions must be made about calculating life expectancy figures for persons or for males and females and about the presentation of confidence intervals. These issues are discussed in detail at the end of this chapter.

**1. Mortality data**

There is a legal obligation for every death in England or Wales to be registered within five days of the person dying. As a result mortality data is both accurate and complete, at least in terms of the basic demographic information such as the number of deaths, the age, the sex and the place of death.

The Office of National Statistics (ONS) releases both monthly and annual mortality files. The annual files contain basic demographic details that are adequate for most analysis, including life expectancy calculations. The monthly files contain additional data fields such as the Communal Establishment code which records deaths that occurred in hospitals, nursing homes or residential homes. However, monthly files must be used with caution; some deaths, for example those that are referred to the coroner can be delayed before appearing on the mortality records. As a result a compilation of monthly mortality files can be judged to provide a complete record of mortality for any one time period only after many subsequent files have been received and checked for deaths from the period of interest.

The postcode of the usual place of residence can be used to allocate deaths to electoral wards, local authorities or Primary Care Trusts (PCTs) using the ‘postcode to ward’ lookup files published by the Organisational Codes Service of the NHS Information Agency. It is worth noting that a nursing home (or other communal establishment) is only recorded as the usual place of residence if ‘the place where the deceased person normally resided previously could no longer be regarded as their usual address for example, because they had lived in the hotel or hospital for a long time, because they were a resident employee there or because they had no more permanent residence elsewhere.’ As a result patients who have recently moved to a nursing home will be recorded as living at their former place of residence. This peculiarity is convenient for most mortality analysis, though care must
be taken to ensure that the population data has been allocated in a similar way.

Age at death is provided on the mortality file and can be used to group the population into the appropriate age bands for life expectancy calculations. If life expectancy is being calculated by gender, the ‘sex’ field can be used to divide the deaths by gender.

Mortality files contain information on the cause of death. Up to eight contributing causes can be recorded, with one flagged as the underlying cause. Whilst this data is not needed for conventional life expectancy calculations, it does enable the calculation of years of life lost by disease specific life expectancy. Information on the deceased person’s occupation, provided by the key information at the time of death registration, is used to derive social class and may be of interest to an epidemiologist analysing variations in mortality. Such information must be handled with more care than the basic fields as it is known to be not quite as accurate.

2. Population data
In order to calculate abridged life tables for small areas, population data broken down by age and sometimes by sex needs to be obtained. This causes most problems for this type of calculation.

2.1 Census derived estimates: problems for LE calculations
The Census is acknowledged to be the gold standard for UK population data and is readily available for areas smaller than electoral wards. A number of problems occur when using the Census data for LE calculations in small areas. The most important difficulty is that the data only produce ward level figures once every ten years, after the national decennial Census. Thus, the further away from a Census year, the less accurate the data, so ward level Census populations can only be reliably used to calculate life expectancy for the period close to a Census year.

2.1.1 Data for the intercensal period
For larger areas, such as local authorities and health regions, ONS produces annual mid-year population estimates in 5-year age bands. Yet even at this level, as the revisions to the earlier published estimates following the release of 2001 Census data show, the mid-year population estimates are not always entirely accurate.

In an attempt to produce a methodology for calculating mid-year population estimates at electoral ward level, Oxford University’s Department of Social Policy published in 2001 the results of a project to estimate 1,998 ward populations. They used a variety of sources including data from child benefits, the Labour Force Survey, and state retirement and incapacity benefits, and constrained the total population of wards in a district to the district mid-year estimate produced by the ONS. These figures were produced primarily for use in the DETR Index of Multiple Deprivation 2000. As the figures were only available in three age bands (under 16s, 16-59s, and 60 and over),
they could not have been used
directly for life expectancy
calculations even if they had
been repeated.\textsuperscript{2}

Until recently mid-year estimates at
ward level have been deemed too
unreliable for regular calculations.
The ONS has now established the
Small Area Population Estimates
(SAPE) project group which is
investigating potential methodologies
to make ward level mid-year estimates
available by age and sex. A paper was
published in 2003 assessing the
various data sources available at local
authority level\textsuperscript{3} and the group will
report in 2004 on the practicalities of
producing these estimates regularly.

In some areas a local authority may
conduct their own local census or
create other population estimates, for
example the Enhanced Electoral
Registration survey in Hampshire.\textsuperscript{4}
Due to the expense involved in such
studies, much like the national
Census data, these are unlikely to be
regularly updated. Some local
authorities produce mid-year
estimates for electoral wards but there
is no consistent methodology.
Comparing such locally produced
figures across larger areas could be
problematic due to the variety of
approaches used.

The London Research Centre is one
body that regularly produces mid-year
estimates for electoral wards, but only
for the London boroughs. Their
method is based upon births and
deaths for populations aged under 15,
and the rest of the population is aged
for each year following the Census.
The International Passenger Survey is
also used allowing adjustments to be
made to the figures to compensate
for population changes as a result
of migration.

\textbf{2.1.2 Other disadvantages of
Census data}

A second disadvantage of Census data
is that, unlike mortality data, it is not
available down to the level of the
individual. Data is already aggregated
into age bands or geographical areas;
the most detailed unit available is the
Enumeration District (ED, up to
1991) or Output Area (OA, from
2001). These consist of groups of
households (approximately 125 in
2001) and, whilst the ONS produces
‘postcode to ED/OA’ and ‘ED/OA to
ward’ lookup files, the data cannot be
easily reordered into new
administrative boundaries when these
occur. Problems can therefore occur
when calculating trends over time for
small areas.

Thirdly, Census data is produced
centrally and then distributed to
users. Although it is readily available
at minimal or no cost to those in the
public sector, there can be delays in
production which cannot be
controlled by the end user. For
example the electoral wards data with
full age breakdown files collected in
April 2001 was not made available
until October 2003 despite being
scheduled many months earlier.

\textbf{2.2 General practice population
register (Exeter) data}

The Exeter database is a record of all
patients registered with a GP practice
in a particular area. There are 98
interlinked Exeter systems across the
country, one for each former Family
Health Service Authority. The register
is active in that the records of those
patients who move in and out of the area or who die are added and removed on a daily basis. It provides a potential source of population data that is readily available and free of charge to analysts within the NHS.

2.2.1 Benefits of Exeter data
Exeter data has a number of advantages over Census data. As a ‘live’ database a snapshot of data can be downloaded at any time. Data for each individual on the register includes their age and postcode so it can easily be sorted into new age bands or new administrative boundaries as required. Because the Exeter register is available as raw data inside the NHS and it can be sorted in-house, there should be minimal delay between the request and the receipt. Another benefit is that the Exeter database is available for the whole of Great Britain so population estimates can be derived nationwide from the same source.

2.2.2 Disadvantages of Exeter data as a source of population estimates
It is impossible to retrospectively extract data from the Exeter database. Data from previous years is only available if it had been downloaded and stored at the time. This has not been routinely done in many areas.

In previous years the number of incomplete records in the Exeter database was thought to be a problem. The majority of these incomplete records were due to inaccuracies in the recorded postcode which meant that patients could not be allocated to administrative areas when the data was analysed. However over the last decade or so a large effort has been made to improve the accuracy of Exeter data. Table 1 shows the improvement in the Exeter data between 1991 and 1997 for all 98 Exeter systems nationwide. By 2000 the recording of postcodes on the Exeter database had improved even further. Taking the West Sussex area as an example, of the 788,639 patients registered on 31st December 2000 all but three could be matched to wards using the NHS February 2001 lookup file. The improvement in Exeter data has been encouraged by increased use. Initially simply a system to hold patient records, the Exeter system now provides data that is used to pay GPs and to monitor progress against national targets for vaccination and screening. It seems likely that the quality of Exeter data will continue to improve as more use is made of it, especially in light of changes resulting from the new GMS contract in April 2004.

<table>
<thead>
<tr>
<th>Year</th>
<th>% of records with incomplete records</th>
<th>% of records with invalid postcodes</th>
<th>% of records with no postcodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>3.2%</td>
<td>0.5%</td>
<td>2.7%</td>
</tr>
<tr>
<td>1994</td>
<td>2.3%</td>
<td>0.5%</td>
<td>1.8%</td>
</tr>
<tr>
<td>1997</td>
<td>1.2%</td>
<td>0.2%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

Source: ONS Population Trends 96. 1991 figures are based on 94 of the 98 FHSA records.
Another disadvantage of the Exeter database for estimating population is that it does not provide complete coverage; anyone not registered with a GP will not be included on the database. Those who rely entirely on private healthcare, long stay patients in a psychiatric institution, long-term prisoners and personnel from the armed forces are consequently excluded.

Despite this, Exeter data is acknowledged to give inflated estimates of the true population. The ‘Estimating with Confidence’ project published figures in 1998 showing that the number of patients registered on the Exeter database has exceeded the number of residents in every year since 1961. This discrepancy is probably caused by delays in removing patients who die or who emigrate and patients who are registered in two or more areas; students are thought to contribute heavily to the latter group. With the increased use of Exeter data and resulting improvements in data quality, it is hoped that this overestimation will diminish.

2.3 Differences between Exeter and Census population data
In 1997 nearly 4 per cent more patients were registered on FHSA registers in England and Wales than were reported in the 1997 mid-year population estimates. In 2003 the ONS reported that Exeter database had 4.2 per cent more patients registered in April 2001 than were recorded in the Census of that year. Although this may suggest the Exeter register has become less suitable as a population count, since the papers were published the ONS has revised the 2001 Census figures upwards and the 1997 mid-year estimates downwards which may counter this apparent trend.

Of the 4.2 per cent overall difference the variation in the male population was 6.1 per cent whilst the female population differed by just 2.4 per cent. In the age category the biggest differences were seen in the 16-44 age group, where Exeter data recorded 8.7 per cent more males and 4.3 per cent more females than the Census 2001. The difference between the Exeter and Census data is also known to vary between regions. Disparity between Census and Exeter Local Authority total populations ranged from -7.3 per cent to +30.1 per cent in 2001.

2.4 Differences between Exeter and Census data for ward populations
It is likely that the disparity between Exeter and Census data would be greater between wards than between local authority populations. Haynes et al looked at data for Norfolk and Suffolk wards and found the 1991 Exeter population to be highest compared to the 1991 Census in urban wards with high population density. In 5 per cent of wards the total Exeter population was found to be over 20 per cent higher than the 1991 Census, although it was within 5 per cent of the Census figure for around half the wards. The population of the 15-44 age group was found to be most inflated on the Exeter database especially for males. The study also revealed an overestimation for the population aged 74+ and an under recording for those aged 0-4 years.
Creasey and Edwards found a similar level of difference at ward level using 1997 data for Hampshire and noted that the presence of military or student populations in a ward could create very large differences. Patient counts by sex were within 5 per cent of the population recorded by local census for 75 per cent of wards. The 45–64 age band was found to suffer from list inflation in all areas, as was the population aged over 75 years. They also concluded that the Exeter database underestimated the population under 1 year of age. Exeter and local census estimates were found to be most similar in the 1–15 and 65–74 age bands.11

Recently the ONS SAPE team compared population data derived from 2001 Exeter data estimates for all wards in England and Wales. The mean percentage difference between Exeter and Census total ward populations was 4.2 per cent. This work has shown a close correlation between the two counts ($r^2 = 0.991$), although only after adjusting the Exeter data to compensate for the exclusion of the prison population.12 Analysis of these differences was not available by age and sex at the time of writing but the same team found that at local authority level Exeter data was closest to the 2001 mid-year estimates for 0-15 year olds of both sexes and for females aged 45. The largest list inflation was found in the 16-44 age group for both sexes, although it was higher for males than for females. The ONS also plans to compare ward figures with specially produced mid-year 2001 ward estimates to compensate for the acknowledged problems with 2001 Census counts in some areas.13

Much of the overestimation in younger age bands may be linked to student populations; Exeter data for the wards in Southampton where many students reside appears inflated in the younger age bands, but those without student populations have list deflation for this group.11 The differences between the figures for males and females are thought to be due to the fact that males take longer to register with a new GP when they change address, they take longer to inform their GP they are moving from area, and they are less likely to have details of their previous GP on registration. For these reasons there is likely to be a delay in removing them from the Exeter system in their former area of residence. Young women, on the other hand, access the health services more frequently and so are more likely to register with a local practice when they move away. In addition screening targets for women, monitored through the Exeter database give a strong incentive to keep records up to date for young females; no GP or PCT wants to have their screening rate brought down by patients who no longer live in the area.

Military populations do not register with GPs and so do not appear on the Exeter database. At county or district level these individuals are likely to have a small effect on total population numbers, but at the ward level there is potential for a far greater impact.

Despite these discrepancies, Exeter data is used by the ONS to gauge internal migration and some areas have used it as one source of information for their own population.
estimates. Haynes et al\textsuperscript{15} concluded that Exeter data was an acceptable alternative to the Census for population estimation at ward level. The Hampshire study concluded that Exeter data appeared to be a potential source of small area population estimates but that comparison with the 2001 Census would provide the real test.

2.5 South East data: Exeter and Census population comparisons
This project focused on the differences between Exeter and Census data in the South East region. Exeter downloads on or around 29th April 2001 were obtained for Hampshire and the Isle of Wight, Kent and Medway, Oxfordshire, Surrey and Sussex giving a total of 1,242 wards (according to the ward boundaries as used in the 2001 Census Standard Tables). Census counts were used as a comparison for all areas (i.e. not the adjusted 2001 mid-year estimates which are only available for local authorities).

When comparing Exeter data for wards, only the population from the appropriate Exeter box was used. For example patients on the West Sussex system who had a Surrey postcode were not included in this analysis. This practice reduced the risk of double counting (in case the Exeter extract received from one area already included ‘fringe patients’ from another area). As a result findings from this analysis replicate those that would be generated from a download of the local Exeter system only, as per routine work within most NHS organisations.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Total ward populations April 2001: comparison of Exeter and Census derived figures for 1242 wards in South East England}
\end{figure}
2.5.1 Total population by electoral ward

Overall there was a close correlation ($r^2 = 0.967$) between total ward populations derived from both Census and Exeter sources, as shown by the scatter diagram in figure 1. The line of best fit suggests that the Exeter data tended to be 5.3 per cent higher than the corresponding Census count for that ward. This data shows a slightly less good fit than the ONS SAPE study for all wards in England and Wales suggesting there may be more variance in the South East than nationally.

In the wards analysed, the mean population difference between Exeter populations and Census counts was +4.7 per cent. Of the 1,242 wards 34 (2.7 per cent) show differences greater than ±20 per cent and 13 (1.0 per cent) have differences between the two population figures of greater than ±30 per cent.

The 13 wards with differences between the two population figures of ±30 per cent and over are shown in Table 2. Where Exeter figures exceed Census figures all the wards have a university based in or near to them. Carfax, Holywell, St Margaret’s and Oxford North are all central Oxford wards with a number of colleges and large numbers of students, whilst the Headington Hill and Northway ward is the site of Oxford Brookes University. The University of Kent’s main campus falls within the Blean Forest ward, the University of Surrey is in the Onslow ward in Guildford and the universities of both Sussex and Brighton are located within the Hollingdean and Stanmer ward. Brunel University’s Runnymede Campus is in Englefield Green.

All of the wards where Exeter data provides a lower population than the Census count are home to a communal establishment that could explain the large difference. The Parkhurst ward has three prisons (Parkhurst, Camp Hill and Albany).

<table>
<thead>
<tr>
<th>Ward</th>
<th>Census population</th>
<th>Exeter population</th>
<th>Difference population</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holywell</td>
<td>4313</td>
<td>10380</td>
<td>6067</td>
<td>141%</td>
</tr>
<tr>
<td>Carfax</td>
<td>4570</td>
<td>10782</td>
<td>6212</td>
<td>136%</td>
</tr>
<tr>
<td>Blean Forest</td>
<td>4674</td>
<td>10846</td>
<td>6172</td>
<td>132%</td>
</tr>
<tr>
<td>Oxford North</td>
<td>5474</td>
<td>8842</td>
<td>3368</td>
<td>62%</td>
</tr>
<tr>
<td>Onslow</td>
<td>8180</td>
<td>12857</td>
<td>4677</td>
<td>57%</td>
</tr>
<tr>
<td>St. Margaret’s</td>
<td>4598</td>
<td>6835</td>
<td>2237</td>
<td>49%</td>
</tr>
<tr>
<td>Hollingdean and Stanmer</td>
<td>14451</td>
<td>21305</td>
<td>6854</td>
<td>47%</td>
</tr>
<tr>
<td>Englefield Green East</td>
<td>5692</td>
<td>7540</td>
<td>1848</td>
<td>32%</td>
</tr>
<tr>
<td>Headington Hill and Northway</td>
<td>4887</td>
<td>6357</td>
<td>1470</td>
<td>30%</td>
</tr>
<tr>
<td>Parkhurst</td>
<td>4434</td>
<td>3090</td>
<td>-1344</td>
<td>-30%</td>
</tr>
<tr>
<td>Brize Norton and Shilton</td>
<td>2742</td>
<td>1797</td>
<td>-945</td>
<td>-34%</td>
</tr>
<tr>
<td>Boarhunt and Southwick</td>
<td>1909</td>
<td>1233</td>
<td>-676</td>
<td>-35%</td>
</tr>
<tr>
<td>Pirbright</td>
<td>4209</td>
<td>2005</td>
<td>-2204</td>
<td>-52%</td>
</tr>
</tbody>
</table>
The findings of previous analysis that suggests there is a greater overestimation in the male population are therefore confirmed. When the differences are explored by age as well, similar patterns to those reported by other studies are again observed. Figure 2 shows the mean difference between the number of registered patients and the resident population for the 1,242 wards in the South East by both age and sex.

As can be seen, inflation on the patient register is more severe for males than females throughout the majority of age bands from age 20 upwards. Compared to Census figures, the Exeter data is highest in the 20-44 age group. Another peak is evident in the 85+ bracket where mean ward differences are approximately 4.3 per cent for females and 9.0 per cent for males. The data also suggests that the under recording reported by Haynes et al in the 0-4 age band occurs primarily in those aged less than 1 year, confirming the findings of Creasey and Edwards. The Exeter data for the South East population aged under 1 year was below the Census data by an average of 8.2 per cent for both sexes, whilst the 1-4 years age group was on average 1.3 per cent over the Census count.

2.5.3 Population for the first year of life

Given the evidence that the Exeter database underestimates the population for the first year of life, the numbers for children under 1 year on the West Sussex Exeter system and the number of live births recorded in the same year (2000-2001) were compared. It was found that the Exeter population aged
under 1 year was 7,060, whilst the Vital Statistics (VS) live births record showed that 7,904 babies had been born to mothers in West Sussex, almost 12 per cent more. Within the county, 94 out of 155 wards had 10 per cent or more live births than were recorded on the Exeter database wards, and 6 wards had over twice the number. It appears that the registration of newborns can sometimes be delayed, resulting in under recording.

It is particularly important that the population aged less than 1 year is entered accurately because the life expectancy model places more weight on young deaths and death rates tend to be high in the first year of life. Substituting the population aged 0 years for the number of VS births in the preceding 12 months would minimise the risk of this occurring. The Government Actuary’s Department’s definitive national life expectancy figures for the UK uses this substitution. Their figures are calculated by replacing the Census-based mid-year population estimates for those aged 0 years with the number of live births in the first year of life.

2.5.4 Looking at individual ward differences

Although the mean difference between the two populations for any age/sex group is around 20 per cent this disguises much greater variation at the level of individual wards. Figure 3 shows the distribution of all differences by age for the 1,242 wards. The maximum difference in any one ward in one age band is 1,000 per cent (in the Holywell ward in Oxford), and on 45 occasions the
difference within one age band is over 100 per cent. However, the bulk of these large differences lie in the 15–54 age group where few deaths occur so it was anticipated that even though large, the differences would have a relatively small impact on life expectancy calculations.

2.6 The impact of population differences on life expectancy figures
The Exeter data does not provide a perfect data source, especially where detail on the age and sex breakdown of a population is required, yet this does not mean necessarily that it is
unsuitable for life expectancy calculations. To assess the data’s suitability as a denominator for LE calculation, LE figures were calculated for all 1,242 wards using both Exeter and Census populations and deaths from 1997 - 2001. The two sets of figures were then compared.

Figure 4 shows the distribution of differences in LE when calculated with the two populations. Although the biggest differences seen are +7.0 years (Carfax) and −5.0 years (Whitely), the average absolute difference for the 1,242 wards was 0.53 years. 1,086 (87.4 per cent) of wards had differences of less than ± one year between LE figures calculated with the Exeter and Census data, and 734 (59.1 per cent) had differences of less than ± 0.5 years. Only 21 (1.7 per cent) of the wards had a difference of over 2 years.

It is possible to analyse which age bands are most responsible for the differences in LE figures when switching between the two sources of population data. By looking at the correlation between the differences in LE and the differences in population by age band it is possible to determine where the differences in population have the greatest impact on LE.

If all wards are included in such analysis, the outlying wards tend to disrupt the calculation of the correlation coefficient. Figure 5 gives one example of this in the 85+ age band. The coefficient of determination ($r^2$) is very low although the majority of results lie close to the line of best fit. Removal of the two outliers raises the coefficient of determination to over 0.5.

Figure 6 shows the coefficients of determination from regression of the differences in LE compared with the differences in population for all wards by age band. Results are presented when the 5 per cent and 0.5 per cent of wards with the largest absolute differences between the Exeter and Census data (within that particular age band) are excluded. As
CHAPTER 4
Life Expectancy Report 2005

Figure 6: Coefficient of determination from regression of LE differences and population differences by age band: 1242 wards in the South East, and with outlying values excluded

The figure shows, when the wards with very large differences between the Exeter and Census data are removed from the analysis, the correlation coefficients reveal that for the vast majority of wards it is the differences in the 85+ age band that have the biggest impact on LE. For most wards around 50 per cent of the differences in life expectancy can be attributed to the differences between the Exeter and Census population for the 85+ age band.

2.7 Conclusions on the use of Exeter data
Although these findings are encouraging, at the present time Exeter data cannot be recommended as a population source for LE calculations unless further work is carried out in order to adjust for the known areas of error.

For most wards Exeter data is likely to provide a close approximation in total numbers if adjustments are made to take into account large prison or student populations. In terms of LE calculations such an adjustment would only be required where very large differences occur, in general the young and middle age groups have little impact on life expectancy because of their small number of deaths.

It is more difficult to make adjustments to compensate for the differences in the final age band where the majority of deaths occur and where much of the difference in the LE calculations with the two populations arises. We recommend that Exeter data is not used as a population denominator unless appropriate adjustments are made.

3. Population life years at risk and standard error
The work of Eayres and Williams has shown that, using their recommended model, a minimum population size of 5,000 person years
at risk should be employed in order to keep the standard error down to ± 2 years. For the majority of UK wards this can be achieved by using 5-years of mortality data because over 99 per cent of the wards have a population of at least 1,000 individuals (Census 2001). Figure 7 shows the standard error values for the life expectancy figures for all wards in the South East using Census 2001 population data and mortality data from 1998-02. The line of best-fit shows that the average SE for a population of 1,000 (i.e. 5,000 population years at risk) is approximately 2 years.

Although 5-years of mortality data must be used, trend data can still be obtained by rolling forward population and mortality data as required. For example, if life expectancy is calculated for 1998-02 using the 2001 Exeter population, the next year these could be updated with 1999-03 mortality data and the 2002 Exeter population. This process is not dissimilar to that used by the GAD in their yearly interim life tables.

4. LE by male and female vs LE by persons
National life tables and the ONS publication of LE at local authority level are produced for males and females. The national data shows that there is a gap of around 4.5 years between males and females. At local authority level there is a variation in the size of the difference between men and women. In 2000 these varied from 0.4 years to 7.5 years but with 95 per cent of local authorities falling between 6.9 and 2.8 years. The more affluent local authorities with higher life expectancies tend to have smaller differences between the sexes.

It could be argued that it is incorrect to combine the measures as it is known that differences exist between life expectancy figures for males and
females and that the magnitude of this variation varies from place to place. For example an area might appear to have a higher than expected life expectancy for persons as a result of a skewed gender ratio. In areas that have unusually high or low life expectancy figures, it would not be possible to tell whether these were the result of good health in the male or female population.

A key issue in electoral ward level LE calculations is the size of the population. The model produced for this project model has shown that a minimum population size of 5,000 person years at risk should be employed in order to maintain an acceptable standard error. As the majority of UK wards have a population of at least 1,000 individuals (Census 2001), by using 5-years of mortality data LE can be calculated for 99 per cent of electoral wards. However if these ward populations were separated by gender, only 88 per cent of UK wards would meet the criteria (that is, have a population of 2,000 persons or more) and it may be necessary to use either 10 years of mortality data for ward level LE calculations (this is not recommended) or to combine wards into larger population units.

Furthermore, presenting data for both males and females may make the interpretation difficult (and this difficulty is magnified if confidence intervals are calculated for both sexes). Life expectancy figures will often be used to provide a summary of the health status of wards within an area. If figures are to be published for males and females, a single list has to be replaced by two. Moreover, to determine which wards have the worst or best health status overall, male and female life expectancies must be combined in some way. If this is to occur anyway, statistically it is more sensible to combine population and mortality rates at the point of calculation rather than to combine the final figures (possibly weighting by the overall proportion of males to females or undertaking some other rough approximation).

Although the problems of combining sexes must be considered, we recommend that for small areas it is only necessary to calculate single life expectancy giving the value for all persons.

5. Confidence intervals
As most LE calculations have been undertaken on national populations, it has not been routine to produce confidence intervals. The life tables produced by the GAD, which are based on the entire population broken down by each year of life, do not contain confidence limits. The ONS has started to produce confidence intervals for LE calculations of local authorities. Eayres and Williams set out the methodology of calculating confidence intervals on small populations, demonstrating that confidence intervals can be accurately calculated for populations larger than 5,000 person years at risk.

The question that arises from this work is whether it is sensible to present confidence intervals for LE calculations at ward level. The argument in favour is that such
calculations allow the user to gauge quickly the likelihood of a perceived difference between two life expectancy figures being the result of the random nature of mortality and the small numbers involved at ward level. The argument against producing confidence intervals for LEs in small areas is that it means producing three figures for each ward resulting in data which are more difficult to present and more difficult to interpret.

Confidence intervals are often used as a simple, visual, test of statistical significance. If two 95% confidence intervals overlap it is 95 per cent certain (19 out of 20) that there is no statistical evidence of a difference between the two estimates. On the other hand, if the confidence limits do not overlap, there is likely to be a significant difference.

Where confidence intervals are used, it is questionable whether it is necessary to use 95% confidence limits; that is, do we need to be so certain of the measure? This degree of confidence was initially used for clinical trials where false conclusions could be literally a matter of life and death. When interpreting life expectancy figures a false conclusion would not have such serious effects, so a 75% confidence limit or a 68% confidence limit (which would be equal to the Standard Error) could be considered acceptable. Either of these limits could be used to identify those areas where life expectancy was ‘more than likely to be different’ from that of another area or from that of the wider population. After all, what is the danger of investigating or even acting on a difference that is not real as opposed to not acting because the difference cannot be statistically demonstrated?

6. Conclusions
Mortality data causes a few problems for the calculation of life expectancy but it is the choice of population data that has a big impact on the figures produced. Census ward populations and the forthcoming ward level mid-year estimates must be assumed to offer the most accurate population denominators for life expectancy calculations. At the present time the Exeter data appears to be too unreliable for this purpose primarily due to the differences between Exeter and Census data in the final (85+) age band.

Using 5-years of population data will result in an acceptable standard error for most wards (around ± 2 years) when LE figures are calculated for persons, as recommended by this project. If separate figures are required for males and females, mortality data for up to 10 years may be needed to maintain standard error values of this magnitude. We do not recommend combining 10 years data because of the changes in the population denominator over that period.

Confidence intervals can be accurately calculated for small populations and can aid the interpretation of LE figures for small areas. On the other hand they add unnecessary complexity when one of the main advantages of the LE measure is that a lay audience can easily grasp the basic concept and use the resulting figures. Where confidence intervals are used consideration should be given to the degree of confidence that is required. The automatic assumption that a 95% confidence interval is required should be questioned and use of a suitable alternative, such as using the standard error as a 68% confidence limit, should be considered.
References

1 Personal communication with ONS General Section. Information from The Handbook for Registration Officers – not publicly available.
4 Creasey R and Edwards R. Can Family Health Authority patient registers supersede local Censuses as the most reliable source of intercensal small area population estimates? Paper presented at the British Society for Population Studies Conference 1999, University College Dublin. Pg. 2
7 Ibid Scott and Kilbey.
8 Ibid SAPE (03)4
9 Ibid SAPE (03)4
11 Ibid Creasey R and Edwards R.
14 ibid Creasey et al prt 1
15 Creasey R and Edwards R. Can Family Health Authority patient registers supersede local Censuses as the most reliable source of intercensal small area population estimates? Approaching the answer Paper presented at the British Society for Population Studies Conference 2000, Netherlands Graduate School of Housing and Urban Research.
16 Ibid Bates A.
The Impact of Nursing Homes

1. Background

The purpose of this chapter is to examine the impact of nursing home deaths on life expectancy (LE) calculations. Our analysis for this part of the project concentrated on West Sussex data where data on nursing homes was readily available. The essence of our findings has been published in a paper in the European Journal of Epidemiology and Community Health.

One of the main uses of LE is as a measure of health status. The drive to produce life expectancy figures for small areas has been fuelled by the need to identify health inequalities at the local level. For this reason it is important that LE figures should not be distorted by confounding factors such as the underlying distribution of nursing homes.

It is well known that residents of nursing homes experience higher death rates than elderly people of the same age living at home. As the distribution of nursing homes is uneven across geographic boundaries, it is possible that the migration of the frail elderly to areas with nursing homes may decrease the life expectancy in those areas. In larger populations, such as local authorities, this effect is likely to be less apparent as the variation in the distribution of nursing homes is unlikely to be so great.

2. History and literature

A number of studies have examined mortality rates within nursing homes and compared them to the general population. Raines and Wight studied patients admitted to nursing homes in Wakefield. They reported standardised mortality ratios (SMRs) of 546 for those admitted from home, and 606 for those admitted to a nursing home from hospital, compared to the population of England and Wales (100) aged 65 and over.

Dale et al looked at mortality in 32 homes in the Manchester area and found five had SMRs of 700 or over compared to all homes in the study, demonstrating a wide variation in mortality rates between homes.

Prior to their dissolution in 2002, Health Authorities had a legal obligation to monitor nursing home deaths. Despite this there appears to be little data available to analyse these deaths. West Surrey Health Authority, however, was one of the few authorities to undertake a comprehensive annual census of nursing home residents and keep an accurate record of nursing home deaths. For the period October 2000 to September 2001 the SMR for nursing home residents compared to the West Surrey population as a whole was 298. Higher ratios of around 400 had been observed in previous years.

In 1995 Williams et al recognised that increased mortality in nursing homes was a potential source of bias in the calculation of SMRs for electoral wards in Croydon. When nursing home deaths were included in their analysis two wards appeared to have significantly higher mortality rates than the rest. With the nursing home deaths excluded these two wards fell into line with the regional average.

More recently, whilst investigating nursing home mortality in 2002, Professor Paul Congdon concluded...
that in areas where over 20 per cent of deaths occurred in nursing homes, there was a real risk that mortality rates could be affected by the location of these establishments.\textsuperscript{6}

The possibility of nursing homes disrupting patterns of mortality at the local level is occasionally acknowledged. In their technical guide to the life expectancy inequalities target, The Department of Health concluded that: ‘It would be possible, in practice, to calculate life expectancy data at the ward level but the figures may not be sufficiently meaningful or robust. The average population in a ward consists of 5,000 people and is therefore strongly influenced by factors such as the number of old people’s homes in the ward.’\textsuperscript{7}

In a recent paper examining the potential for monitoring GP death rates, Aylin et al also acknowledged that patients in nursing homes were one of a number of case mix factors which could increase mortality rates and therefore increase the risk of triggering false alarms.\textsuperscript{8} Scanlon confirmed this trend for East Sussex: ‘We have analysed practice level mortality data for general practices in East Sussex. The data confirms the view that practices treating more deprived populations and practices delivering services to nursing homes and hospices have higher mortality rates compared with other practices.’\textsuperscript{9} More recently Mohamed et al demonstrated that the high mortality associated with two general practitioners correlated closely with the number of deaths that occurred in nursing homes.\textsuperscript{10}

In September 2003 the ONS report ‘Life expectancy at birth: methodology options for small populations’ noted that the presence of nursing homes in an electoral ward may lead to local migration effects which will influence mortality rates.\textsuperscript{11}

Two studies from Canada have examined the impact of nursing home deaths on life expectancy in greater detail than these British studies. Veugelers and Hornibrook (2002) looked at life expectancies for neighbourhoods of 2,500 to 41,000 persons (average population 14,677) in Nova Scotia, Canada. They were able to assign back to their former area of residence individuals who died in nursing homes, and so were able to calculate life expectancies both with and without this redistribution. Their results showed differences of up to two years in the resulting life expectancies.\textsuperscript{12}

Manuel et al (1998) calculated LE for the 42 public health units in Ontario and commented on the impact of migration to nursing homes by examining the probability of death in younger and older age groups. Unfortunately the methodology of the paper is weak. There is no information on the number or proportion of nursing home deaths. The investigators simply looked at a correlation between deaths in the 10-24 and 70-74 age groups. They concluded that because health units with a high death rate in the elderly age group also had a high death rate in the younger age group, there is no evidence of a nursing home impact.\textsuperscript{13}
3. Nursing home deaths

In order to investigate the impact of deaths in nursing homes on LE calculations some measure of ‘nursing homes’ is needed. There can be significant changes in the number of nursing homes and beds in an area from year to year and, as small area mortality calculations tend to use deaths over a period of time, ideally this measure should span a similar period. Unfortunately few sources of nursing home data can be obtained retrospectively and where this is possible records often only date back a couple of years. It is rare that data can be obtained to cover the full five years required for small area life expectancy calculations.

In West Sussex, nursing home deaths had been monitored over a period of time, and accurate figures for the proportion of nursing home deaths by electoral ward were available for the period 1997-01. Unfortunately it was not possible for this study to identify nursing home deaths (NHD) for all wards in the South East with the same degree of accuracy due to the lack of a comprehensive list of nursing homes open during the study period.

4. Impact of nursing home deaths on LE in West Sussex and Croydon

The 155 electoral wards in West Sussex County (2001 ward boundaries) were used to examine the impact of nursing home deaths on LE calculations in small areas for the 5-year period 1997-2001. For this analysis the proportion of all deaths, where the deceased person resided in a nursing home, was calculated. For simplicity this measure will be referred to as nursing home deaths (NHDs).

NHDs were calculated for West Sussex wards using a locally compiled list of all nursing homes open in West Sussex over this period of time. Data were also available from Croydon PCT, so the 27 wards in Croydon have been included for comparison in parts of this discussion.

West Sussex has around 120 nursing homes with about 4,200 beds. At the time of the 2001 Census, 2,924 West Sussex residents were living in nursing homes, distributed unevenly amongst the seven local authorities. An analysis of the NHDs within West Sussex wards confirms that deaths are...
predominantly amongst the elderly (figure 1) with a mean age of 81.5 years and a standard deviation of 12.9 years. The distribution of NHDs varies considerably between wards: in a third of wards there were no deaths of residents, but in 5 wards over 50 per cent of deaths were in nursing home residents (figure 2).

4.1 Nursing home deaths and deprivation

The relationship between LE and deprivation is well documented. Deprived areas tend to have higher mortality rates than their affluent equivalents and consequently lower life expectancies. Figure 3 shows the strong correlation between LE figures (1998-02) for wards in the South East and the number of people receiving income support in 2001, as a proportion of the ward population.

If nursing homes tend to be distributed in deprived or affluent areas, the observed relationship between NHDs and LE could be a result of the confounding effect of deprivation. This was tested by examining the correlation between deprivation and the proportion of nursing home deaths. Table 1 confirms that for the 155 West Sussex...
wards there appears to be no significant relationship between deprivation and NHDs for a number of common deprivation indices.

This trend does not appear to be exclusive to West Sussex; nursing home deaths in Croydon wards do not show a significant relationship with deprivation indicators either.

4.2 LE and proportion of nursing home deaths

Figure 4 shows a comparison of the distribution of life expectancy figures for electoral wards in West Sussex with nursing home deaths against those without. The distributions of the two sets of wards follow a similar shape but the life expectancy figures for wards with nursing home deaths are approximately two years lower.

A statistically significant association between LE and NHDs is confirmed by linear regression. The relationship is truly linear as the residuals around the line of best fit exhibit a near normal distribution. The proportion of nursing home deaths can explain approximately 37 per cent of the variation in life expectancy among West Sussex wards. This relationship appears to be fairly constant across areas; a similar line of best fit would be drawn for the same variables for Croydon PCT although the overall relationship is weaker. It can be estimated that, in wards where 25 per cent of those dying are nursing home residents, calculations of life expectancy at birth will be approximately two years below similar wards without nursing homes (figure 5).

When life expectancy is calculated at ages other than birth, the correlation with nursing homes is much stronger. Figure 6 shows the relationship between LE at 65 and NHDs for all wards in West Sussex. This pattern can be explained: when calculating life expectancy at older ages we are
only concerned with older deaths. Because nursing home residents make up a greater proportion of older deaths, they will have a bigger impact on LE calculated at older ages.

4.3 Relationship between LE, deprivation and nursing home deaths
The independent impact of deprivation and nursing homes can be confirmed by multiple linear regression. Table 2 presents the results of multiple linear regression analysis between life expectancy, measures of deprivation and nursing home deaths (NHDs) for West Sussex wards. Weighted multiple linear regression was used, with residuals weighted by the size of the ward population. This technique allows more weight to be given to data for larger wards, which tend to have a smaller standard error and thus a more reliable estimate of life expectancy.

The figures in table 2 demonstrate that both nursing homes and deprivation contribute towards life
Together the two variables can explain between 60 and 70 per cent of the variation in life expectancy figures in West Sussex, depending on the deprivation indicator used. The standardised coefficient shows the relative importance of the two independent variables, namely deprivation and NHDs. The similarity of these coefficients confirms that each variable explains a similar proportion of the variation in life expectancy. These findings provide evidence for the link between nursing homes and life expectancy and confirm nursing home deaths to be a suitable measure for investigating this relationship.

5. Nursing homes and LE in the South East

In the absence of accurate data on nursing homes throughout the South East region, it has not been possible to examine the relationship between LE and nursing homes for all areas in this study. Yet if this relationship is ignored there is a danger that life expectancy figures for small areas could be wrongly interpreted.

To provide a starting point for local analysis, the LE figures for the South East available through SEPHO have been published along with two approximate measures of nursing homes: firstly, the number of deaths where the deceased was resident in a nursing home as proportion of all deaths 65 and over; secondly, the number of nursing home residents as proportion of the total population aged 65 and over. The first measure is derived from both the public health mortality files and the ONS communal mortality file. The second measure is derived from the Census. Both are weak measures as they are derived from a ‘snapshot’ of nursing homes rather than being based on the true nursing home population throughout the entire study period (see appendix B for further details on these measures).

Our analysis suggests that these measures of nursing homes should be treated with caution. For example, when comparing both with LE for wards in West Sussex NHDs could explain only 15% and 16% of the variation in LE between wards. The impact of nursing homes on life expectancy figures locally can only be fully understood when an accurate measure of nursing homes spanning the study period has been produced.

6. Nursing homes and residential care homes

Nursing homes are not the only institutions known to attract elderly migrants. Frail elderly people often move across administrative borders to residential care homes. Although elderly migrants to residential care homes are likely to be in better health than those moving to nursing homes, it
is possible that migration to these establishments may also impact on LE figures for small areas.

In practice it can prove difficult to distinguish between nursing and residential home beds for the purpose of defining a nursing home variable. A number of homes are now dual registered and provide both residential and nursing accommodation. The division of beds between these two specialities within such homes may not be either clear or consistent. In addition, a number of homes may change designation over time from nursing to residential or vice versa. This can pose problems in mortality analysis where we may want data over a number of years.

We recommend that, when conducting further analysis in this area, the impact of residential care homes as well as nursing homes on life expectancy figures for small areas is investigated.

7. Effects on SMRs

The Standardised Mortality Ratio (SMR) is widely used to describe geographic variations in mortality and have also been used in resource allocation, with areas with higher mortality rates attracting more resources. Although individual SMRs cannot be compared directly with each other (and are expressed as ratios rather than in units of years) they are often used in preference to life expectancy in small area analysis due to the relatively small size of the associated standard error figures. SMRs for wards are often included in Director of Public Health Annual Reports (for example, East Sussex, 2003\textsuperscript{14}) and post-Shipman attempts to monitor deaths rates for GPs’ practices have also made use of them.\textsuperscript{15}

As a measure the SMR is more susceptible to elderly migration than life expectancy as it weights all deaths equally. One extra infant death will carry the same weight as one additional elderly death. In contrast, in the life table model used to calculate life expectancy, a younger death will have more impact than an older death, as the younger death involves more years of life lost. It is reasonable therefore to assume that SMRs will be affected more by elderly migration than LE will.
To examine the above hypothesis, the all-cause SMR for wards in West Sussex was compared with the proportion of nursing home deaths. As shown in figure 7, the correlation was far stronger than that found between LE and nursing home deaths. The proportion of nursing home deaths can explain 57 per cent of the variation in SMRs between wards.

This finding suggests that care should be taken when interpreting SMRs in small areas in populations that have a significant number of nursing homes. When calculating SMRs, it is important to understand the impact of nursing home deaths.

8. Means of adjusting for nursing homes

Various ways of removing the nursing home bias from life expectancy calculations have been explored during this project. An ideal scenario would involve allocating the deaths of those dying in nursing homes and the nursing home population back to the geographical regions where they lived before moving to the home. This was achieved by Veugelers and Hornibrook with Canadian mortality data, but would be far from straightforward within the UK as previous addresses are not recorded on the mortality database.

8.1 Life expectancy to 75

Another option would be to publish life expectancies solely up to the age of 75. This would exclude the deaths of the elderly, and thus the nursing home bias, from the analysis but it would also reduce the sensitivity of the measure. The figures produced would be affected primarily by deaths in the very young and would not respond to health inequalities in the old. With the health of the elderly becoming an increasingly important topic, this is a major disadvantage for a summary measure of health status.

Another disadvantage of this approach is that the life expectancy figures calculated only to age 75 begin to differ from what is perceived as length of life. For example, values for West Sussex would range from 66 to 73 years. The complexity of describing this measure to a lay audience and explaining why life expectancy appears so low may outweigh some of the advantage of calculating it in this way.

8.2 Removal of nursing home patients and deaths from the analysis

The effects of elderly migration could be minimised by removing both the nursing home population and the corresponding deaths from each ward. However this could artificially boost all ward level life expectancies (by removing the population with high death rates) and, as discussed in the appendix, it is not easy to get precise numbers for the nursing home population over a period of time.

An adjustment of this nature is likely to involve removing the precise number of nursing home deaths from each year within the analysis and then removing the population as counted in one year only. This leaves the adjustment open to the potential of a numerator-denominator bias. The same criticism applies to most small area mortality calculations (population estimates from one year are often used alongside up to five years of mortality data), but it is of heightened importance here due to the small numbers in ward nursing home populations.
8.3 Regression based adjustment
Another possible way of removing the nursing home bias involves adjusting life expectancy figures according to the gradient of the line of best fit between life expectancy and nursing home deaths. All wards with nursing home deaths would be raised by the product of the gradient of the regression line and the proportion of nursing home deaths. In West Sussex this would result in adding two years onto life expectancy figures for those wards with 25 per cent of nursing home deaths. A ward with 50 per cent nursing home deaths would have its life expectancy boosted by approximately four years.

Such an adjustment will remove any bias due to nursing homes (there will be no correlation between nursing home deaths and the resulting distribution) but it does have disadvantages, for instance assuming the impact of nursing homes on ward life expectancy is uniform over a large area.

The accuracy of this adjustment is very hard to verify. In addition, the confidence intervals which can be calculated for life expectancy figures would be redundant because the possible error from the crude adjustment is likely to be far larger than the error from the LE calculation.

The biggest criticism of this adjustment is that ward life expectancies are being increased for wards with nursing homes but are not being decreased in those areas without nursing homes as would happen if these deaths could be reassigned back to their former ward of residence. Unfortunately it is not a simple operation to redistribute the ‘excess’ mortality back to those wards with no nursing homes.

9. Conclusions
LE is often used as a proxy measure of health status and as a tool for describing inequalities in health, so it is important that the data are not affected by local differences in the population density of nursing home residents.

Our research has demonstrated that nursing home deaths can have a strong impact on LE figures for small populations. Indeed, LE figures for electoral wards in West Sussex showed as strong a correlation with nursing homes as with deprivation, while these two variables showed no relationship with each other. We recommend, therefore, that the effect of nursing homes must be taken into account before life expectancy figures for small areas are published or used in decision-making. The same caution applies to other measures such as SMRs when used for small populations.

The most appropriate way of presenting LE figures for small areas is to present data on the proportion of nursing home deaths together with LE figures. Analysts who calculate LEs at the local level should spend some time identifying nursing home deaths from the public health mortality file, using local knowledge of the distribution of nursing homes within their PCT.

To facilitate further research in this area the deaths of nursing home residents should be made more easily identifiable and routinely monitored. There is potential for these deaths to be marked on UK mortality files.
References


Calculating life expectancy

This section contains a more detailed description of the methods used for the calculation of life expectancy (LE). There is some overlap between this section and chapter three in this report in order that each chapter can be read independently.

1. The current life table: the Chiang method

Table A1 shows an entire abridged current life table, using the mortality data for males in England over the period 1998-00. The table has been completed using the methods described by Chiang. This methodology is used by The Office for National Statistics to produce life expectancies at Health Authority and Local Authority level. It has also been used by the London Public Health Observatory in their study of the health inequality targets as applied to London. A description of each of the terms used in the life table is given below.

1.1 Age interval

The life table is broken down by age intervals. Each age interval is denoted by the subscript \( i \) where \( i = 0 \) to \( \omega \). The age at the start of age interval \( i \) is denoted by \( x_i \).

Table A1

| Age Interval | Age At Start Of Interval | Age Interval Width | Fraction of Last Age | Population Years At Risk | Number Of Deaths In Interval | Annual Death Rate In Interval | Probability Of Dying In Interval | Probability Of Surviving Interval | Number Alive At Start Of Interval | Number Dying In Interval | Number Of Years Lived Beyond Start Of Interval | Total Number Of Years Lived Beyond Start Of Interval | Observed Expectation Of Life At Start Of Interval |
|--------------|--------------------------|--------------------|----------------------|--------------------------|----------------------------|-----------------------------|-------------------------------|---------------------------------|-----------------------------|-------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| \( i \)     | \( x_i \)                | \( x_{i+1} \)      | \( n_i \)            | \( q_i \)                | \( p_i \)                  | \( M_i \)                   | \( q_i \)                      | \( p_i \)                       | \( l_i \)                     | \( d_i \)                      | \( L_i \)                                    | \( L_i + T_i \)                                  | \( e_i \)                                      |
| 0            | 0                        | 0                  | 0.1                  | 894,146                  | 5,651                      | 0.00632                     | 0.00628                       | 0.99372                        | 100,000                      | 628                         | 99,434                                     | 7,544,321                                    | 75.44                                         |
| 1            | 1                        | 1-4                | 0.5                  | 3,711,609                | 1,071                      | 0.00029                     | 0.00115                       | 0.99885                        | 99,372                       | 115                         | 397,257                                    | 7,444,886                                   | 74.92                                         |
| 2            | 5                        | 5-9                | 0.5                  | 5,103,517                | 654                        | 0.00013                     | 0.00064                       | 0.99936                        | 99,257                       | 64                          | 496,126                                    | 7,047,629                                   | 71.00                                         |
| 3            | 10                       | 10-14              | 0.5                  | 4,788,565                | 781                        | 0.00016                     | 0.00082                       | 0.99918                        | 99,193                       | 81                          | 495,765                                    | 6,551,504                                   | 66.05                                         |
| 4            | 15                       | 15-19              | 0.5                  | 4,793,794                | 2,527                      | 0.00053                     | 0.00263                       | 0.99737                        | 99,113                       | 261                         | 494,910                                    | 6,055,739                                   | 61.10                                         |
| 5            | 20                       | 20-24              | 0.5                  | 4,643,359                | 3,627                      | 0.00078                     | 0.00390                       | 0.99610                        | 98,852                       | 385                         | 493,295                                    | 5,560,829                                   | 56.25                                         |
| 6            | 25                       | 25-29              | 0.5                  | 5,620,487                | 4,921                      | 0.00088                     | 0.00437                       | 0.99563                        | 98,466                       | 430                         | 491,256                                    | 5,067,534                                   | 51.46                                         |
| 7            | 30                       | 30-34              | 0.5                  | 6,245,244                | 6,139                      | 0.00098                     | 0.00490                       | 0.99510                        | 98,036                       | 481                         | 488,979                                    | 4,576,277                                   | 46.68                                         |
| 8            | 35                       | 35-39              | 0.5                  | 5,988,450                | 7,363                      | 0.00123                     | 0.00613                       | 0.99387                        | 97,556                       | 598                         | 486,283                                    | 4,087,298                                   | 41.90                                         |
| 9            | 40                       | 40-44              | 0.5                  | 5,076,082                | 9,372                      | 0.00185                     | 0.00919                       | 0.99081                        | 96,958                       | 891                         | 482,561                                    | 3,601,015                                   | 37.14                                         |
| 10           | 45                       | 45-49              | 0.5                  | 4,735,027                | 14,206                     | 0.00300                     | 0.01489                       | 0.98511                        | 96,067                       | 1,430                        | 476,757                                    | 3,118,455                                   | 32.46                                         |
| 11           | 50                       | 50-54              | 0.5                  | 4,981,283                | 23,696                     | 0.00476                     | 0.02351                       | 0.97649                        | 94,636                       | 2,224                        | 467,620                                    | 2,641,697                                   | 27.91                                         |
| 12           | 55                       | 55-59              | 0.5                  | 3,878,763                | 31,702                     | 0.00817                     | 0.04005                       | 0.95995                        | 92,412                       | 3,701                        | 452,807                                    | 2,174,077                                   | 23.53                                         |
| 13           | 60                       | 60-64              | 0.5                  | 3,466,624                | 47,754                     | 0.01378                     | 0.06658                       | 0.93342                        | 88,711                       | 5,907                        | 428,788                                    | 1,721,270                                   | 19.40                                         |
| 14           | 65                       | 65-69              | 0.5                  | 3,074,475                | 72,157                     | 0.02347                     | 0.11084                       | 0.88916                        | 82,804                       | 9,178                        | 391,075                                    | 1,292,482                                   | 15.61                                         |
| 15           | 70                       | 70-74              | 0.5                  | 2,626,525                | 105,230                    | 0.04006                     | 0.18208                       | 0.81792                        | 73,626                       | 13,406                       | 334,614                                    | 901,407                                     | 12.24                                         |
| 16           | 75                       | 75-79              | 0.5                  | 2,075,460                | 135,473                    | 0.06527                     | 0.28058                       | 0.71942                        | 60,220                       | 60,220                      | 258,857                                    | 566,793                                     | 9.41                                          |
| 17           | 80                       | 80-84              | 0.5                  | 1,088,004                | 114,357                    | 0.10511                     | 0.41618                       | 0.58382                        | 43,323                       | 18,030                      | 171,540                                    | 307,936                                     | 7.11                                          |
| 18           | 85                       | 85+                | 10.8                 | 782,553                  | 145,116                    | 0.18544                     | 1.00000                       | 0.00000                        | 25,293                       | 25,293                      | 136,395                                    | 136,395                                     | 5.39                                          |
Age intervals can then be defined as:

\[ x_i \leq x < x_{i+1} \quad \text{for } i = 0, 1, \ldots, \omega - 1 \]

\[ x_i \leq x \quad \text{for } i = \omega \]

(the final age interval is open ended, i.e. there is no upper age limit).

1.2 Interval width
The width in years of age interval \( i \) is denoted by \( n_i \).

1.3 Fraction of last age interval of life
Members of the hypothetical cohort who die during an age interval do not all do so at either the beginning or the end of the interval but at various points throughout its length. The fraction of last age interval of life is the average fraction of the age interval that they survive before dying. Denoted by \( a_i \).

Generally the deaths are considered to be evenly distributed across the age interval, giving a fraction of 0.5. For the under 1 age group however it is known that deaths are not evenly distributed, being much more likely to occur in the perinatal and neonatal periods, and the fraction is set to 0.1 to reflect this:

\[ a_i = 0.1 \quad \text{for } i = 0; \]

\[ a_i = 0.5 \quad \text{for } i = 1, 2, \ldots, \omega \]

1.4 Population years at risk
The population years at risk for age interval \( i \), denoted by \( P_i \).

1.5 Number of deaths in interval
The number of deaths observed in the study population for age interval \( i \), denoted by \( D_i \).

1.6 Annual death rate in interval
The average annual age-specific mortality rate for age interval \( i \), denoted by \( M_i \):

\[ M_i = D_i / P_i \quad \text{for } i = 0, 1, \ldots, \omega \]

1.7 Probability of dying/surviving in interval
The probability that a member of the hypothetical cohort will die/survive during the age interval \( i \). The probability of dying is denoted by \( q_i \). The probability of surviving is denoted by \( p_i \).

Either probability can be calculated from the mortality rate \( M_i \) but the exact nature of the relationship depends on the methodology used and the assumptions that the chosen methodology makes.

The remainder of the life table can be completed by applying either the probability of dying or the probability of surviving to the hypothetical cohort. The two approaches are equivalent as:

\[ p_i = 1 - q_i \]

For the final age interval the probability of dying is equal to 1 and the probability of surviving is 0:

\[ p_i = 0, q_i = 1 \quad \text{for } i = \omega \]

The Chiang methodology, also known as the linear method, assumes that within an age interval the deaths of members of the hypothetical cohort are evenly distributed. Consider figure A1 for the age interval \( i \) of width \( n_i \).
The number alive at the start of the interval is \( l_i \) and the number of deaths during the interval is \( d_i \). The probability of death \( q_i \) is given by:

\[
q_i = \frac{d_i}{l_i}
\]

The cohort’s average annual mortality rate during the interval \( M_i \) is defined by:

\[
M_i = \frac{d_i}{L_i}
\]

giving:

\[
d_i = M_i L_i
\]

where \( L_i \) is the total years lived by the cohort during the interval and is equivalent to population years at risk.

\( L_i \) is given by:

\[
L_i = n_i (l_i - d_i) + a_i n_i d_i
\]

Rearranging gives:

\[
l_i = \frac{L_i + (1 - a_i) n_i d_i}{n_i}
\]

Substituting for \( d_i \) gives:

\[
q_i = \frac{n_i M_i}{l + (1 - a_i) n_l M_i}
\]

Thus in the life table, \( q_i \) can be calculated from the estimate of \( M_i \) given by the observed number of deaths and population years at risk of the real population: \( M_i = \frac{D_i}{P_i} \).

1.8 Number alive at start of interval

The number of the hypothetical cohort alive at age \( x_i \), the start of age interval \( i \). Denoted by \( l_i \).

The size of the cohort at birth, \( l_0 \) is arbitrary and is usually set to 100,000. The number alive at the start of subsequent age intervals can be calculated by either:

- applying the probability of surviving the previous age interval to the number alive at start of the previous age interval;
- or subtracting the number dying during the previous age interval from the number alive at start of the previous age interval:
1.9 Number dying in interval

The number of the hypothetical cohort dying in the age interval \( i \). Denoted by \( d_i \). The number dying in interval can be calculated by either:

- taking the number alive at start of the age interval and subtracting the number alive at start of the next age interval;
- or applying the probability of dying in the interval to the number alive at start of the interval.

As the final age interval is open-ended all those alive at the start of the interval will die during it:

\[
\begin{align*}
\text{for } i &= 0, 1, \ldots, \omega-1, \\
d_i &= l_i - l_{i+1} \quad \text{OR} \quad d_i = l_i \, q_i \\
\text{for } i &= \omega, \\
d_i &= l_i 
\end{align*}
\]

1.10 Number of years lived in interval

The number of person years lived during the age interval \( i \) by the members of the cohort who are alive at the start of the interval. Denoted by \( L_i \).

Each member of the cohort who survives the age interval \( i \) contributes \( n_i \) years (the width of the interval). Each member of the cohort who dies during the interval contributes \( n_i \) multiplied by \( a_i \) years (the width of the interval multiplied by the average fraction of the interval that they survived before dying).

The above does not work in the final age interval as it is open-ended. Instead it is assumed that survival in this age interval is exponential i.e.

\[
\begin{align*}
l_i &= 100,000 \quad \text{for } i = 0 \\
l_i &= l_{i+1} \, p_{i+1} \quad \text{OR} \\
l_i &= l_{i+1} - d_{i+1} \quad \text{for } i = 1, 2, \ldots, \omega 
\end{align*}
\]

1.11 Total number of years lived beyond start of interval

The total number of person years still to be lived by members of the cohort who are alive at the start of the age interval. The equation is the sum of number of years lived in the age interval and all the subsequent age intervals. Denoted by \( T_i \).

\[
\begin{align*}
T_i &= L_i + L_{i+1} + \ldots + L_\omega \quad \text{OR} \\
T_i &= T_{i+1} + L_i \quad \text{for } i = 0, 1, \ldots, \omega 
\end{align*}
\]

1.12 Observed expectation of life at start of interval

By dividing the total number of years lived beyond the start of interval by the number alive at the start of interval, the average number of years that each of the members of the cohort alive at the start of the interval can expect to live is obtained. This observed expectation of life is denoted by \( e_i \).

\[
\begin{align*}
e_i &= T_i / l_i \quad \text{for } i = 0, 1, \ldots, \omega 
\end{align*}
\]

For this study ‘life expectancy’ is defined as the mean expectation of life at birth, \( e_0 \).
1.13 Calculating the 95% confidence intervals
For the Chiang, or linear, method the variance of the life expectancy, is given by the formula:

\[ S_{e_i}^2 = \frac{1}{F_i} \sum_{j=0}^{t} F_j \{ (1-a_j)n_j + e_{j+1} \} S_{p_j}^2 \]

\[ S_{p_j}^2 = \frac{q_j p_j}{D_j} \]

Assuming that the estimate of life expectancy is normally distributed the 95% confidence intervals are found in the usual way:

\[ e_i \pm 1.96S_{e_i} \]

where \( S_{e_i} \) is the standard error of \( e_i \) given by \( \sqrt{S_{e_i}^2} \).

2. The Government Actuary’s Department (GAD) method
The Government Actuary’s Department produces the official national life tables and life expectancy figures. Every ten years, based on data around a Census year, the Department produces a set of graduated (smoothed) life tables for England & Wales and Scotland. For each year in between ungraduated interim life tables are produced for the UK and its constituent countries on a three-year average basis. The GAD does not produce life tables nor life expectancy figures at a sub-national level.

The GAD life expectancies are produced from complete life tables but the theory is essentially the same as for the abridged tables described above and uses the same assumption that deaths are evenly distributed over the age interval.

For the first age interval under 1 year \((i=0)\) the probability of dying \(q_0\) is calculated directly using infant deaths at <4 weeks, 1-2 months, 3-5 months, 6-8 months and 9-11 months and the corresponding ‘at risk’ population derived for each group from the monthly birth figures.

For the remaining age intervals, the probability of dying during the interval is given by the equation:

\[ q_i = \frac{2M_i}{(2 + M_i)} \]

for \( i = 1,2,\ldots, \omega-1 \)

This is equivalent to the Chiang equation for abridged life tables, given that \( a_1 = \omega\) and \( n_1 = 1.\) For this study the GAD England Life Tables 1998-00 were used to give the reference life expectancies for comparison with those derived from the test methods and their variations.3

3. The Silcocks method
One of the alternative methods for calculating life expectancies is described by Silcocks et al in their paper on the statistical considerations of calculating sub Health Authority level life expectancies.7 This methodology has already been used by the Trent Public Health Observatory to calculate life expectancy at electoral ward level.
Silcocks’s methodology uses abridged life tables but differs from the Chiang method in the way it estimates the probability of survival during an age interval. It assumes that the average annual mortality rate of the cohort during the interval is constant and therefore that survival is exponential. The probability of survival $p_i$ is given by:

$$p_i = \exp(-n_i M_i)$$

where $n_i$ is the width of the interval, and $M_i$ the average annual mortality rate.

As with the Chiang method $M_i$ can be estimated from the observed number of deaths and the population years at risk of the real population:

$$M_i = D_i / P_i$$

For complete life tables this exponential method is less accurate than the linear method as the assumption of constant mortality risk throughout the interval does not reflect real life. In human populations the mortality risk for age intervals after childhood increases slowly with age at a rate consistent to that of the linear method.

When using the exponential method to estimate the probability of survival, it is usual to apply it also to the estimation of the number of years lived in interval. The method described above in the description of the life table is the linear method, which assumes the deaths are evenly distributed across the age interval. If the mortality risk is constant, the assumption will not hold; the number of deaths will be greatest at the start of the interval and decrease throughout the interval as the size of the cohort decreases. The exponential method estimate for $L_i$ is given by:

$$L_i = n_i (l_i - l_{i+1}) \frac{\ln l_i - \ln l_{i+1}}{M_i}$$

However Silcocks et al do not use this estimate but instead return to a linear method using the formula:

$$L_i = n_i (l_i + l_{i+1}) \frac{\ln l_i - \ln l_{i+1}}{2}$$

This is equivalent to the Chiang linear formula for $L_i$ given above but with the proviso that $a_i = 0.5$, i.e. Silcocks et al use $a_i = 0.5$ for all age intervals.

For the final age band $\omega$, however, $l_{\omega+i}$ is equal to zero (because no one survives the final age band) and $n_{\omega}$ (the width of the final age band) is taken to be twice the mean survival in this age group assuming exponential survival, i.e. $2/\bar{M}_{\omega}$. This results in the same formula as used in the Chiang method:

$$L_i = \frac{l_i}{\bar{M}_i}$$

The hypothetical width of the final age interval as used by the Silcocks methodology is reflected in the example life table in Table A1 where the width of the 85+ age interval is given as 10.8 years.
3.1 Calculating the 95% confidence intervals

Silcocks et al derived the following formula for the estimation of the variance of the life expectancy:

\[ S^2_{\epsilon_i} = \frac{1}{F_i} \sum \left( \frac{\partial \epsilon_i}{\partial z_j} \right)^2 S^2_j \]

where \( z_j \) are independent variables given by:

\[ z_j = n_j M_j \quad \text{with} \quad S^2_{z_j} = \frac{n_j D_j}{P_j} \]

for \( j = i, i+1, \ldots, \omega-1 \)

\[ z_\omega = \frac{2}{M_\omega} \quad \text{with} \quad S^2_{z_\omega} = \frac{4}{D_\omega M_\omega} \quad \text{for} \quad j = \omega \]

and

\[ \frac{\partial \epsilon_i}{\partial z_j} = -\left( \frac{n_j}{2} + \epsilon_{j+1} \right) \]

for \( j = i, i+1, \ldots, \omega-1 \)

\[ \frac{\partial \epsilon_\omega}{\partial z_j} = \frac{l_j}{2} \quad \text{for} \quad j = \omega \]

The variances \( S^2_{z_j} \) are based on the assumption that the observed deaths \( D_j \) follow a Poisson distribution and therefore also have a variance of \( D_j \). This differs from the Chiang assumption that the deaths are binomially distributed. In instances where the mortality is low and death is a rare event the two distributions are approximately equivalent.

However for some of the age intervals under consideration the magnitude of the mortality rate is such that death cannot be considered as a rare event. In such circumstances there is some debate as to which distribution is the most appropriate to use. For the purpose of this study each methodology was tested using deaths simulated from the distribution upon which that methodology is based, i.e. the Chiang method was tested using deaths simulated from a binomial distribution and the Silcocks method by deaths from a Poisson distribution.

Another difference from the Chiang method is the inclusion of a term for the variance of the life expectancy associated with the final age band. In the Chiang method the variance of the life expectancy is the weighted sum of the variance of the probability of survival across all the age intervals. For the final age interval the probability of survival is, by definition, zero and has zero variance. However, Silcocks et al argue that in the case of the final age interval the life expectancy is dependent not on the probability of survival but on the mean length of survival \( (1/M_\omega) \). The weighted variance of this length of survival (expressed in their paper in terms of the hypothetical width of the final age band) is then included in the overall variance of the life expectancy. There is no reason why this additional variance term cannot be included in the Chiang methodology, and for this study the effect of its inclusion is investigated.

4. Problems arising from small numbers

One of the aims of this study was to determine the robustness of life expectancy estimates for electoral wards. At this level the populations and numbers of deaths involved are relatively small: the average English electoral ward has a population of around 6,000 persons and experiences about 60 deaths a year. The standard errors of life expectancies calculated from such numbers are likely to be relatively large and the confidence intervals wide. It may be necessary to aggregate electoral wards over a
number of years or into larger geographical areas to increase the population years at risk and thus reduce the standard error of the life expectancy.

5. Problems arising from death counts of zero

Another problem associated with small numbers is the occurrence of zero deaths in an age interval. Silcocks et al noted that such a count gives an estimate of zero for the sample variance of the age interval, which is clearly an underestimation of the true variation. This results in an underestimation in the total life table variance and therefore the life expectancy standard error. The greater the number of zero death counts within the life table the greater the underestimation of the standard error will be.

Silcocks et al suggested two ad-hoc corrections for such zero death counts. The first is 0.693 being the Poisson mean for which the probability of observing zero deaths is 0.5. The second is 3.000 being the Poisson mean for which the probability of observing zero deaths is 0.05 (i.e. the upper 95% confidence limit for the mean number of deaths when zero were observed). To these a third possible correction was added: using the expected number of deaths given the size of the population in the age interval and the appropriate exact England age-specific mortality rate.

Such corrections, if simply placed into the observed deaths column of the life table, will affect the life expectancy estimate itself as well as the standard error. The effect is likely to be significant at the electoral ward level as a) zero deaths will be a common occurrence, particularly for younger age intervals; and b) the fixed substitutions of 0.693 and 3 are relatively large in size. Therefore, in addition to investigating the effects of performing these simple substitutions, the study examines the effect of applying the substitutions only within the standard error calculations. Where the standard error calculations themselves make reference to the age-specific life expectancies $e_i$, then the uncorrected $e_i$ will be used.

A zero death count is particularly important in the final age interval. If there are zero observed deaths, the death rate ($M_0$) is zero and the members of the hypothetical cohort who survive to the start of the final age interval then have an infinite mean length of survival ($1/M_0$), giving an infinite life expectancy. Therefore, in the final age interval the zero death rate is replaced by the death rate ($M_0$) of England. This is done irrespective of whatever substitution method is being employed.

6. Study methodology

The approach taken in this study was similar to that employed by Silcocks et al to test their life expectancy and standard error formulae. Monte Carlo simulations were performed to calculate the life expectancy and its standard error for a hypothetical electoral ward population experiencing given underlying age-specific mortality rates.
A single simulation consists of a large number of repetitions. For each repetition the underlying age-specific mortality rates and populations are used to generate a random count of deaths for each age interval from a known probability distribution. These death counts are then inserted into the life table and the resulting life expectancy and standard error recorded. The results from all the repetitions give a distribution of life expectancies and standard errors from which inferences can be made.

In this study:
- The underlying age-specific mortality rates used were the England male rates for 1998-00;
- The hypothetical populations were such that their age structures were the same as those of England males 1998-00;
- Each simulation consisted of 10,000 repetitions;
- Simulations using Chiang methodology assumed the binomial distribution for generating the random counts of death;
- Simulation using Silcocks’s methodology assumed the Poisson distribution for generating the random counts of death.

In effect each simulation asks the question “What is the random distribution of life expectancies that is likely to be seen for an electoral ward with the same age structure and experiencing the same underlying mortality rates as England males?”

To investigate any differences between the Chiang and Silcocks methodologies and the effect of using different age intervals within the life tables, simulations were performed for each of the following life table models:
- Chiang 5-year abridged to 85+
- Chiang 5-year abridged to 90+
- Chiang 5-year abridged to 95+
- Chiang 10-year abridged to 85+
- Chiang 10-year abridged to 90+
- Chiang 10-year abridged to 95+
- Silcocks’s 5-year abridged to 85+
- Silcocks’s 5-year abridged to 90+
- Silcocks’s 5-year abridged to 95+
- Silcocks’s 10-year abridged to 85+
- Silcocks’s 10-year abridged to 95+

To investigate the effect of population size on life expectancy and its standard error, simulations were repeated for the following hypothetical electoral ward population sizes:
- 500
- 1,000
- 5,000
- 10,000
- 25,000
- 50,000

To investigate the potential problem caused by occurrences of zero death counts and the effects of suggested solutions, simulations were repeated using the following substitutions for an occurrence of zero deaths within an age interval:
- No substitution
- Substitution of 0.693 deaths in both life expectancy and standard error calculations
- Substitution of 3 deaths in both life expectancy and standard error calculations
- Substitution of expected number of deaths in both life expectancy and standard error calculations
- Substitution of 0.693 deaths in standard error calculation only
In addition to recording the simulated data, the application also recorded the expected life expectancy and the standard error for each abridged life table model calculated using the underlying mortality rates directly. Two further sheets calculated the expected life expectancy using the underlying mortality rates in unabridged life tables, one using the Chiang methodology and the other Silcocks’s.

7. Results

7.1 Reference LE

Table 2 (chapter 3) shows the LE for males in England 1998-00 calculated by the various life table models. These figures are not the results of simulations but are calculated from the exact age-specific mortality rates. They are the reference figures against which the simulated life expectancies are compared and are referred to hereafter as the ‘reference life expectancy’.

The ‘gold standard’ figure is calculated using the Government Actuary’s Department unabridged life table and is equal to 75.42 years. Both the Chiang and Silcocks’s unabridged life tables give a life expectancy equal to that of the GAD.

The abridged life tables using the Chiang method produced very good approximations to the GAD reference figure, with the largest difference being only 0.04 years in size (for the 10-year 85+ model). As might be expected, the Chiang life tables with 5-year age intervals produced slightly better approximations than those with 10-year intervals.
Each of the abridged life tables using the Silcocks method produced a higher estimate of life expectancy and had a greater difference from the GAD reference figure than its corresponding Chiang life table. This is particularly true for the 10-year 95+ life table. As with the Chiang methodology, the life tables based on 5-year age intervals gave better approximations than those based on 10-year intervals.

7.2 Simulated LEs

Table 3 (chapter 3) shows the mean of the simulated life expectancies for each of the life table models for the hypothetical ward with an all age population ranging from 500 to 50,000. Each of the means is the average of the 10,000 life expectancies generated by the repetitions of the simulation for that particular model.

The hypothetical ward experiences the same underlying mortality rates irrespective of its population size and therefore has the same reference life expectancy for all the populations. The mean of the simulated life expectancies was also expected to be independent of the population size but for the data in the table this is clearly not the case. For all of the life table models it can be seen that smaller population sizes gave higher estimates of the life expectancy. This is illustrated in figure 2 (chapter 3). As the ward population size increases, the means of the simulated life expectancies approach their corresponding reference life expectancy.

It is necessary to understand the reason behind these observations as, if they are genuine, they have a fundamental impact on the appropriateness of using the life expectancy methodologies on small populations. One possibility is that the observations are artefactual and result from poorly performing binomial and Poisson random number generators used to generate the death counts. However, tests of both generators suggest that they both perform as well for small numbers as they do for large and are therefore not responsible. The Office for National Statistics (ONS) has independently reported similar observations in a study using its own binomial random number generators. They suggest that the observations may be a result of the increased frequency of zero death counts within the life tables for the smaller populations but do not outline an explanatory mechanism. It is our opinion that this is not the case.

Any unusual effect of the zero death counts would affect the right hand side of the life expectancy distribution more than the left since the greater the number of zero death counts in a life table, the greater its life expectancy estimate. This would result in a skewed distribution. Consider figure 4 (chapter 3) which shows the distribution of the life expectancies simulated using the Chiang methodology in a 5-year abridged life table to ages 85+ (the life table used by ONS for its sub-national life expectancy estimates). The X and Y scales are the same for all the plots to help illustrate the way the distribution changes with ward population size. It confirms Silcocks et al’s observation that life expectancy is normally distributed, and shows that this is true even for ward populations of 1,000. Although the
mean life expectancy increases as the population becomes smaller, there does not appear to be any evidence of the distribution being skewed.

Another possibility is the effect of ‘grouping’ within the distribution. Grouping occurs as a consequence of smaller numbers of deaths in the life tables giving smaller numbers of permutations of unique possible life expectancies. Grouping affects the right hand side of the distribution more than the left as the life tables with the smaller numbers result in higher life expectancy estimates. In our simulations for the Chiang 5-year abridged life table to 85+ model, the 10,000 repetitions for 50,000 population gave rise to 10,000 unique life expectancies. For 1,000 population this figure fell, but only to 9,108. It then dropped off rapidly for the 500 population to 4,646. Given these figures it seems unlikely that grouping is responsible for the relationship between the mean of the simulated life expectancies and the population size as it does not really become a factor until the population falls below 1,000. Further evidence of this comes from inspecting the median value of the simulated life expectancy estimates. Given that the distribution is normal, if the increase in the mean as the population decreases is explained by the increased grouping in the right hand side of the distribution, then the median life expectancy would be expected to remain constant. However, as Table A2 shows for the Chiang 5-year abridged life table to 85+, this is not the case. The median of the simulated life expectancies, as well as the mean, increases as the population decreases. This suggests that the entire simulated life expectancy distribution is being shifted to the right.

7.3 Choice of methodology for small populations
Further inspection of figure 2 (see chapter 3) shows that for ward populations greater than 5,000 the choice of methodology is the most important factor in estimating the life expectancy, with all of the Chiang models giving better approximations than all of the Silcocks’s models. However, for smaller populations this is no longer necessarily the case. For populations of 500 the most important factor is the choice of final age interval with life table models using 85+ giving the best approximations to the GAD reference figure. This suggests that a significant proportion, if not all, of the ‘drift’ in the mean of the simulated life expectancies for smaller populations is being caused by the final age interval.

7.4 The final age band
In both the Chiang and Silcocks methodology the mean survival within the final age interval of a member of the cohort who survives to its beginning is given by \(1/M_{89}\), where \(M_{89}\) is the observed (or simulated) mortality rate of the final age interval. Figure A2 shows the distribution of 10,000 \(M_{89}\) for the 85+ age interval of a Chiang
Figure A2

Inspection Of The Simulated Mortality Rates In The 85+ Age Interval By Population Size Using Mortality Rates For Males Aged 85+, England 1998-00

Expected Mean Mortality Rate = 0.185 Deaths per Person-Year At Risk

Population: 100,000

Population: 50,000

Population: 25,000

Population: 10,000

Population: 5,000

Population: 1,000

methodology life table model for ward population sizes from 100,000 down to 1,000 (note that the populations refer to the all ages population of the ward, not the population of the final age interval). The distribution of the simulated rates is a weighted binomial (being the binomially distributed death counts divided by the 85+ population). As the population size increases, this distribution tends towards the normal and the standard deviation decreases. The mean of the distribution, which is an estimate of the underlying mortality rate, is independent of the population size.

Figure A3 shows what happens when these simulated rates $M_0$ are transformed to give the simulated mean survival of the final age interval, $1/M_0$. It can be seen that the right hand tail of the survival distribution becomes stretched. The degree of this stretching depends on the standard deviation of the original rate distribution. For a large population where the standard deviation of the rates is small, the corresponding survival distribution remains a close approximation to the normal and its mean gives a good approximation to the expected mean survival as calculated from the exact underlying mortality rate. As the population decreases the standard deviation of the rates increases and the transformed survival distribution becomes more stretched, shifting the distribution mean further and further to the right.

Using 90+ or 95+ as the final age interval instead of 85+ will generally result in larger drifts in the life expectancy estimate for small populations. This is because the populations in these age intervals are smaller, resulting in larger standard deviations in the rates and therefore greater ‘stretching’ of the transformed survival times.

7.5 Zero deaths and the final age band
If the population is small enough, zero deaths may start to occur in the final age interval. In these instances the rate is zero and survival time is infinite. The survival distribution is stretched infinitely to the right and has an infinite mean. To prevent this from happening in the life tables, whenever a zero death count occurs in the final age interval, the exact underlying mortality rate is used instead.

In the plot of the rate distribution for 1,000 population (figure A2) approximately 1,100 of the repetitions had a zero simulated death count. In the corresponding transformed distribution (figure A3) they have been substituted by the exact underlying rate, resulting in an extra 1,100 counts with a survival of 5.39 years and no infinite survivals.

7.6 Life expectancy standard error
For each of the life tables tested there are three measures of the life expectancy standard error. The first is the reference standard error, calculated by the life table formula using the exact underlying mortality rates. The second is the observed standard error found by
Figure A3

Inspection Of The Simulated Mean Survival Time In The 85+ Age Interval By Population Size Using Mortality Rates For Males Aged 85+, England 1998-00

Expected Mean Survival Time = (1/0.185) = 5.39 Years

measuring the standard deviation of the life expectancies calculated by each repetition of the simulated life table. The third is the mean estimated standard error, i.e. the mean of the standard errors calculated by each repetition of the simulated life table.

The first two measures were used in the study by Silcocks et al. They tested the validity of the standard error formula they had derived by comparing the reference standard error that it generated against the observed standard error measured from their simulated life expectancies.

In this study the behaviour of the mean estimated standard error was of particular interest. This is because the estimate of the standard error calculated from a single repetition of a simulated life table is equivalent to the estimate of the standard error that would be obtained from a real electoral ward. By describing the distribution of these simulated standard errors, the way in which the standard error estimated from real data will behave can be inferred. The reference standard error and the observed standard error are included to provide reference figures against which comparisons can be made.

In table 4 (chapter 3) all three measures of the life expectancy standard error are given for each of the life tables models tested and for each of the selected population sizes.

By comparing the unadjusted with the Chiang (Adjusted) methodology, identical observed standard errors are produced. This is unsurprising as both methodologies are identical in the way they calculate the life expectancy.

Where they differ is that the Chiang (Adjusted) methodology includes the variance of the final age interval in the calculation of the standard error. As a result the adjusted methodology generates reference standard errors and mean estimated standard errors that are greater than those produced by the unadjusted method. For populations of 1,000 or greater the adjusted reference standard errors and the mean estimated standard errors compare better against the observed standard errors than the unadjusted. This implies that the adjusted methodology gives the better measure of the true life expectancy standard error.

Other observations from table 4 (chapter 3) are: that the Silcocks methodology gives similar measures of the standard error to the Chiang (Adjusted) methodology; that for populations over 1,000 the width of the age intervals and the choice of the final age interval have little effect on the standard error; and that the life expectancy standard error increases as the population size decreases.

Figure A4 shows how the three measures of the life expectancy standard error for the Chiang (Adjusted) 5-year abridged life table to 85+ vary with the population size. Assuming that the male mortality rates for England 1998-00 apply, a population of 10,000 has a life expectancy standard error of approximately 1.4 years (giving a 95% confidence interval (CI) of approximately ± 2.8 years). For a 5,000 population this increases to 2.0 years (a 95% CI of ± 3.9 years), and for a 1,000 population to 4.3 years (a 95% CI of ± 8.4 years).
7.7 Standard error and zero deaths

A notable feature of figure A4 is the good agreement between the mean estimated standard error and the reference and observed standard errors. This suggests that, as the population decreases, the increasing occurrence of zero death counts in the life table has little impact on standard error.

Zero deaths should only affect the mean estimated standard error. The reference standard error is calculated from a life table that uses the exact underlying mortality rates and therefore has no zero death counts. The observed standard error is not calculated from a formula but is measured from the distribution of simulated life expectancies, the calculation of which is not affected by the occurrence of zero death counts. In figure A4 therefore, the mean estimated standard error would be expected to underestimate the true standard error and to diverge from the two reference standard errors as the population size decreases. This occurs to a small degree for a population size of 500 but it was anticipated that the underestimation would be greater and would be evident at larger population sizes.

Figure A5 shows for the Chiang (Adjusted) 5-year abridged life table to 85+ for selected population sizes, a box-plot of the distribution of the estimated standard errors broken down by the number of zero death counts that occurred in the life table that produced them. In this particular life table model there are 19 age intervals. For a population size of 50,000 the number of these age intervals that contain zero deaths ranges from 0 to 6, with an average of between 2 and 3 in each repetition of the life table. For a population size of 1,000 this has increased to a range of 8 to 18 with an average of between 13 and 14. The box-plots clearly show the effect of zero deaths on the estimated standard error. For a given population size as the number of zero death counts increases, the estimated standard error decreases. This occurs even for a population size of 50,000 and, as the bottom two plots indicate, can also be observed when the Silcocks methodology or 10-year age intervals are used.
Figure A5

Simulated Life Expectancy Standard Error Versus Occurrence Of Zero Deaths In Simulated Life Table By Population Size
Using Chiang (Adjusted) Methodology - 5 Year Abridged Life Table To 85+
Assuming Male Mortality Rates For England 1998-00

There is evidence therefore that within a simulation of a particular life table model, those repetitions with higher numbers of zero death counts give lower estimates of the life expectancy standard error but that the overall mean remains a good estimate of the true standard error. The most probable explanation for this apparent contradiction is similar to that described above to explain the overestimation of the life expectancy. When the age-specific mortality rates are transformed to give their associated variances, the resulting distributions become skewed. The smaller the population, the greater the skew. This results in a mean overestimation of the standard error that counterbalances the underestimation caused by the occurrence of zero deaths.

7.8 Distribution of standard error estimates
A further characteristic of the box-plots in figure A5 is that for small populations the distribution of the estimated standard errors is not normal with the upper tail being longer than the lower tail. This can also be seen in figure A6 which shows histograms of the distribution of the estimated standard errors by population size. For population sizes down to 10,000 the distribution closely follows the normal. For a 5,000 population the right hand tail of the distribution starts to become stretched and for population sizes of 1,000 and 500 the distribution is highly skewed.

The width of the distribution indicates how well the mean estimated standard error is being calculated. For a population size of 50,000 the mean, or best, estimate of the standard error is 0.61 with a standard deviation of 0.03 giving a fairly narrow 95% confidence interval for the estimate of approximately 0.55 to 0.67. For a population size of 5,000 the mean estimate of the standard error is 1.96 but the associated standard deviation has risen to 0.40 giving a relatively wide 95% confidence interval (still assuming a normal approximation) of approximately 1.18 to 2.74.

7.9 Substituting for zero deaths
Table 5 (chapter 3) shows the results of the various substitution methods used in an attempt to compensate for the anticipated effect of zero death counts in the Chiang (Adjusted) 5-year abridged life table to 85+. The first three methods entailed substituting each occurrence of zero deaths in the life table with either 0.693, 3, or the expected number of deaths respectively. After making the substitutions the life expectancy and its standard error were calculated in the normal way. It is clear from the table that making substitutions in this manner adversely affects the life expectancy estimate, causing it to be underestimated. This is particularly true for the fixed substitutions of 0.693 and 3 deaths suggested by Silcocks et al. As the population size decreases the effect of these fixed substitutions becomes proportionately greater, resulting in extremely large underestimation. The last three methods avoid affecting the life expectancy estimate by making the substitutions only in those parts of the life table concerned with calculating the standard error. The effect of all six methods on the mean estimated standard error is illustrated in figure A7. None of the methods tested produced better estimates of the standard error than if no substitutions had been made.
Figure A6

Distribution Of Simulated Life Expectancy Standard Error By Population Size Using Chiang (Adjusted) Methodology - 5 Year Abridged Life Table To 85+
Assuming Male Mortality Rates For England 1998-00

8. Conclusions

The first objective of the study was to compare any differences between the life expectancies estimated by using the Chiang and the Silcocks methodologies. For the population structure and the mortality rates considered, there was good agreement between the estimates produced by the two methodologies. The Chiang methodology gave marginally better estimates than Silcocks’s when compared to the reference estimate produced by the Government Actuary’s Department.

From a methodological point of view the Chiang methodology may be considered to be preferable as it is more consistent in its assumptions. It uses linear assumptions in estimating both the probability of dying and the number of years lived by the cohort during an age interval. The Silcocks methodology uses an exponential assumption to estimate the probability of survival but a linear assumption to estimate the years of life lived. It is a somewhat pedantic point and the Silcocks methodology is by no means incorrect in doing this, merely a little inconsistent. The Chiang methodology is the more widely used and is employed by the Office of National Statistics (ONS) to produce Health Authority and Local Authority level life expectancy estimates. No advantage is gained by using the Silcocks methodology so it is recommended that the Chiang methodology should be used to calculate life expectancy for smaller populations. For estimating the life expectancy standard error, however, we recommend that the Chiang methodology be adjusted to include the variance associated with the final age band, i.e. that the Chiang (Adjusted) methodology be used.

This study showed that for large populations the choice of the age interval width and of the final age interval had little effect on the estimate of life expectancy or its standard error. For small populations however, the choice of the final age interval becomes increasingly important. Life expectancy becomes increasingly overestimated as the population size diminishes. This effect is greatest when a final age interval of 95+ is used and least when 85+ is used. Currently ONS uses 5-year age intervals to 85+ in its life table model. They are considering the possibility of using 90+ or 95+ for larger populations (Local
Authority level and higher) but retaining 85+ for any potential calculations for smaller populations. Given the overestimation effect for small populations, we strongly recommend that for small populations 85+ be used as the final age interval. The choice of a width of 5-years or 10-years for the other age intervals is more arbitrary and we recommend that 5-years is used to retain consistency with the ONS methods used for larger populations.

Investigation of the robustness of the life expectancy estimates for small populations showed that the estimates are normally distributed and that this remains true even for very small populations. However, as expected the standard error of the life expectancy estimates increases as the population size decreases. A population of 5,000 with the same age structure and experiencing the same mortality rates as England males in 1998-00 would have a life expectancy standard error of approximately 2 years, giving a 95% confidence interval for its life expectancy estimate of ±4 years. For a population of 1,000 this rises to over ±8 years. At a Local Authority level the difference between the highest and lowest male life expectancies in England in the period 2001-03 was approximately 8.5 years. Although this differential is expected to be greater for smaller populations, it is clear that for population sizes below 5,000 it will be difficult to show that statistically significant differences exist between different populations, except when comparing those from opposite extremes of the range.

The estimate of the standard error is itself subject to sampling variation and this variation increases as the population size falls. The 95% confidence interval of the estimated life expectancy standard error quoted above for a population of 5,000 is itself ±0.8 years, showing that a population of this size may not give a good estimate of the true standard error. A further problem with estimating the life expectancy standard error of small populations is that its distribution, whilst being normal for large populations, becomes increasingly skewed for smaller ones, particularly for population sizes below 5,000.

Given these problems of overestimation of the life expectancy, its increasingly wide confidence intervals and the increasingly poor estimation of its standard error, we recommend that life expectancy calculations should not be performed for populations less than 5,000 in size. It should be noted that by ‘population’ we actually mean person years at risk. For example, if a person life expectancy is required for a population of 2,500 at least two years’ aggregated data should be used; if the life expectancy of a specific gender is required then this should be doubled to at least four years.

Our investigations also showed that the problem of increasing numbers of zero death counts within the life tables of small populations does exist but that it did not affect the estimate of the life expectancy standard error to the degree anticipated. It is our belief that a mechanism similar to the one which causes the overestimation of the life expectancy of small populations causes an overestimation of the standard error, counterbalancing any underestimation resulting from the zero death counts.
Various methods of correcting the standard error estimate by substituting zero death counts were tested but none were found to give better results than the simple uncorrected method. We recommend therefore that when calculating life expectancy for a small population no substitution be made for an occurrence of a zero death count. The one exception to this is where the zero death count occurs in the final age interval. Such a count, if not substituted, would lead to an infinite life expectancy. In such instances we recommend that the appropriate national age specific mortality rate be used for the final age interval.

9. Summary of recommendations
For calculating life expectancy at electoral ward level we recommend:

- Life expectancy should not be calculated for a population with less than 5,000 person-years at risk;
- The life expectancy and its standard error should be estimated from an abridged life table with 5-year age intervals (0, 1-4, 5-9...) and a final age interval of 85+ using the Chiang (Adjusted) methodology;
- Where a life table contains a death count of zero in the 85+ age interval the appropriate national mortality rate should be used;
- Where a life table contains a death count of zero in an age interval other than 85+ this zero count should not be corrected or substituted.
References


Counting nursing home deaths

In this appendix a number of sources of data that can be used in analysing the impact of nursing homes patients on LE are presented. There are a number of potential variables relating to the count of nursing homes and the choice of which to use may depend on local data availability and suitability. There are three main types of variables that can be considered:

- The number of registered beds
- The number of residents in nursing homes
- The number of deaths in nursing homes

1. The number of nursing homes in a ward

The number of registered homes in a geographical area can be obtained from the National Care Standards Commission (NCSC) which maintains a list of nursing homes together with the postcode of the home, so it is possible to allocate these homes to an electoral ward. Unfortunately the data only date back to April 2002 when the NCSC took over the task of monitoring nursing homes from local authorities and former health authorities.

A list of all communal establishments in England and Wales is also available from the ONS, via the NHS Information Authority website. This list gives the postcode and address for each establishment but does not provide information on the type of institution, making it impossible to distinguish between nursing homes and other institutions such as hospitals or university halls of residence. ONS are able to provide an additional field for this file at a fee, but again these data only date back to 2002.

2. The number of nursing home residents or beds

The NCSC list contains data on the number of beds registered at each in each home. However, the occupancy rate of nursing homes varies substantially between homes in one area, between regions, and over time. All available nursing home beds are unlikely to be filled for the entire study period. Conversely, more than one person may occupy any one particular nursing home bed over the course of a year. For example, if a resident leaves or dies, another may move in to replace them. Whilst this still equates to ‘one bed year’, simply counting the number of beds in such a situation could still underestimate the impact of certain homes.

Another measure of nursing home residents can be obtained from the Census. This gives a measure of the number of Communal Establishment (CE) residents and the proportion of residents in nursing homes. The data is available to the level of Census Output Areas and, combined with the population for the geography in question, this data can be used to determine the proportion of the resident population who live in nursing homes. The data has the added advantage of being readily available across England and Wales in an identical form.

However Census data has its problems. It is only available every ten years and ward level data from 1991 is not directly comparable with that from 2001 due to changing ward boundaries. In addition, the Census only provides a snapshot of the nursing home population on the day the Census is conducted.
3. Identifying individuals in nursing homes using Exeter and mortality data

Two other sources can be used for the nursing home variable: Exeter data and mortality records. Both include the residential address with postcode for all individuals contained in the files. If the addresses and postcodes are known for all nursing homes in the study area, it is possible to check the residential postcode for each person on the GP list or mortality file against the list of nursing home postcodes. Where postcodes match, those individuals can be marked for further attention. Many of these postcodes include other residential properties in addition to a nursing home and so, in order that individuals are not wrongly identified as nursing home residents, the text of the address needs to be checked to confirm that it is a nursing home and not a neighbouring address.

The greatest barrier to this method is trying to obtain a complete list of nursing homes. The lists held by the NCSC or ONS can be used but both suffer the aforementioned problem of only covering a short period of time. In some areas it may be possible to obtain local lists of nursing homes that can be compiled to cover the entire study period. The former West Sussex Health Authority maintained a list of all nursing homes that operated in the county up to 2001. Similar data is available for Croydon and West Surrey and it is likely that comparable sources can be found for other areas.

It is possible to use a combination of methods to ensure all nursing home residents on a mortality file or on the Exeter database are identified. The NCSC list can be used in conjunction with the ONS list, locally compiled lists, and an additional search to identify the word ‘nursing’ within the text of the address lines can be undertaken.

Some areas routinely follow this procedure. Croydon PCT for example maintains a list of local nursing homes and uses this to mark all nursing home patients on both the Exeter and mortality data so they can be easily identified on either file. Figure B1 shows that these two data sources produce closely related, although not identical, results.

4. Nursing home population and deaths of nursing home residents

Identifying individuals from the Exeter database gives a measure of the nursing home population. Conceptually this measure is easily understood and compares directly

Figure B1: Comparing proportions of nursing home deaths (PHMF 1998-2002) and nursing home residents (Exeter April 2001) for Croydon wards

\[ R^2 = 0.8067 \]
with other similar indicators (such as the Census nursing home population or the count of registered beds). Data obtained from the Exeter database is still time specific and so only offers a snapshot of the nursing home population, unless regular Exeter downloads are available to track residents over time. This would allow an average nursing home population to be calculated appropriate to the period of study.

The use of mortality data to determine the nursing home variable has a number of advantages: the records are accurate and complete; the deaths of nursing home residents can be identified for the whole period of study; the data being used as a numerator is exactly the same as in any mortality calculations; and the measure of nursing home deaths rather than one of population means the data also compensates for the different mortality rates between homes.

The studies in West Surrey where these deaths were monitored showed that death rates could vary considerably between different homes. In a given year some homes might have death rates of approaching twice the average for all nursing homes, whilst others may have no deaths and therefore a death rate of zero. Some of this is random variation but other aspects of this variation could be a result of the type of home or the type of patient being admitted.

As discussed in chapter four, individuals who have recently moved to nursing homes may still have their former address marked as a place of residence on the mortality files. This tends to be useful in mortality analysis as any impact of these elderly migrants to nursing homes is removed. If the former addresses are available for all nursing home deaths, it may be possible to redistribute them back to the former geography of residence and remove any nursing home bias, as was done in the Veugelers study discussed in Chapter five.

5. Identifying Nursing home deaths using mortality files

Nursing home deaths can be identified from both monthly mortality returns and the shortened annual summary as both contain the postcode and address for place of residence which is required for this process. An additional code for communal establishments is provided on the monthly files, when the death occurred in such an institution. However, it is the code for the usual place of residence rather than the place of death which is really required. As nursing home residents may die in a hospital and non-nursing home residents may move to a nursing home shortly before their death, using the code for place of death risks causing discrepancies between population and mortality data.

As the monthly mortality files already code the place of death, it should require little additional work for the ONS to provide a code for the place of residence for those individuals living in communal establishments. Such a change would greatly facilitate identification of nursing home patients on mortality files. The process outlined in section 3.3 can identify the deaths of nursing home residents. To make meaningful comparisons between geographical areas once the number of deaths is obtained, the number of nursing home deaths can be expressed as a proportion of all deaths in the area.
# Reader Information

<table>
<thead>
<tr>
<th>Document purpose:</th>
<th>Sharing information on calculating small area life expectancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Technical report – Calculating Life Expectancy in small areas</td>
</tr>
<tr>
<td>Authors:</td>
<td>Ted Williams, Hywell Dinsdale, Daniel Eayres, Farhang Tahzib</td>
</tr>
<tr>
<td>Acknowledgments:</td>
<td>SEPHO agreed to provide funding for the project and a steering group was set up to guide development and implementation. Membership included Dr Ted Williams (WSHA), Dr Farhang Tahzib (WSHA), Dr Anita Sims (SEPHO), Graham Evans (WSHA), Daniel Eayres (National Centre for Health Outcomes Development), Myer Glickman (Office of National Statistics), Keith Hardman (West Sussex County Council), Victoria Axell and Natasha De Meric. Additional thanks to the National Portrait Gallery, Professor Chiang, David Pinder (Woking PCT) and David Osbourne (Croydon PCT).</td>
</tr>
<tr>
<td>Publication date:</td>
<td>November 2005</td>
</tr>
<tr>
<td>Target Audience:</td>
<td>PCT, NHS and Local Authority colleagues with an interest in small area mortality indicators</td>
</tr>
<tr>
<td>Description:</td>
<td>A report which covers the history, methodological options and practical problems relating to calculating small area life expectancy</td>
</tr>
</tbody>
</table>
| Contact details:  | South East Public Health Observatory  
4150 Chancellor Court  
Oxford Business Park South  
Oxford OX4 2GX  
Tel: 01865 334714  
www.sepho.org.uk  

Directorate of Public Health  
Adur, Arun and Worthing Teaching PCT  
The Causeway  
Worthing  
BN12 6BT  
Tel: 01903 708400  
www.healthinaaw.nhs.uk |