National Statistics Methodological Series No.33

Life expectancy at birth: methodological options for small populations

Barbara Toson
Allan Baker
Office for National Statistics

ISBN 1 85774 552 3

Contact points

For enquires about this publication, contact the Editor, Samantha Higgs
Tel: 01329 813660
E-mail: samantha.higgs@ons.gov.uk

To order this publication, call 01329 813660

For general enquires, contact the National Statistics Customer Enquiry Centre on **0845 601 3034**
(minicom: 01633 812399)
E-mail: info@statistics.gov.uk
Fax: 01633 652747
Letters: Customer Enquiry Centre , Room D115, Government Buildings, Cardiff Road
Newport NP10 8XG

You can also find National Statistics on the Internet – go to www.statistics.gov.uk

About the Methodology Series

The aim of this series is to publish monographs with a substantial methodological interest written by people across the Government Statistical Service. Findings can be included provided that they illustrate methodology, but the series is not for findings *per se*.

Contributors should send their submissions to the series editor at the address below. Before doing so, however, they should clear the work with their line manager. The submissions will then be subjected to peer review, external or internal, or both, as the series editor see fit. Authors can submit their work for peer reviewed journals as well.

The series is aimed at getting out results quickly and easily. It is the intention that authors will prepare the final documents themselves, on their own PCs. The style has therefore been kept simple. Style guidelines specifying the fonts, type size, margin settings, etc. are available from the series editor (at the address below). The series editor should need to do minimal editing.

A National Statistics publication

National Statistics are produced to high professional standards set out in the National Statistics Code of Practice. They undergo regular quality assurance reviews to ensure that they meet customer needs. They are produced free from any political interference.

Navigate through this document by using Bookmarks, Thumbnails, or Links from the Contents listing below.
There are more sub-categories in the Bookmarks than there are in the Contents listing.
Prevent the printing of these instructions by unchecking 'Annotations' in the print dialogue box.

Contents

Acknowledgements

The authors would like to thank all those who provided very helpful comments on early drafts of this report:

Andrew Hinde – University of Southampton
Iain Currie – Heriot-Watt University, Edinburgh
Adrian Gallop – Government Actuary's Department
Daniel Eayres – National Centre for Health Outcomes Development
Justine Fitzpatrick – London Health Observatory
Steve Smallwood – ONS
Clare Griffiths – ONS
Myer Glickman – ONS

John Charlton from Methodology Group at ONS also advised on the use of Monte Carlo simulations to test methodological options, and has been valuable in approving the methodological conclusions reached.

David Jenner (East Midlands Public Health Observatory) kindly supplied a copy of the life table used for comparison purposes in Section 3.

1. Introduction

Life expectancy at birth has been used as a measure of the health status of the population of England and Wales since the 1840s. It was employed in some of the earliest reports of the Registrar General to illustrate the great differences in mortality experienced by populations in different parts of the country. For example, in 1841 life expectancy at birth for men in Surrey was 44 years, compared to 25 years for men in Liverpool.[1]

This tradition of using life expectancy at birth as an indicator of geographic inequalities in health has been continued by the Office for National Statistics (ONS) in recent years, with the publication of results for health and local authorities in the United Kingdom. The calculation of national life expectancy results for the United Kingdom and its constituent countries remains the responsibility of the Government Actuary's Department.

Initiating recent work at ONS on this issue, Griffiths and Fitzpatrick selected life expectancy as an indicator to examine geographic inequalities in mortality in the United Kingdom in 1995–1997.[2] They stated that life expectancy was chosen as '…it is a summary measure of mortality at every age that allows comparisons to be made between areas and time periods without the need to assume a particular standard population.' Thus results are comparable for areas which may have very diverse population structures.

Griffiths and Fitzpatrick reported a difference of ten years between the local authorities with the highest and lowest male life expectancy at birth – 78.4 years in Chiltern, 68.4 years in Glasgow City. ONS has continued to report on life expectancy at birth in the United Kingdom, partly because of its intrinsic merit as a means of analysing geographic variations in mortality, but also to inform one of the two national health inequality targets announced by the Department of Health in 2001.[3] The aim of this target is to reduce by at least 10 per cent the gap between the fifth of local authorities with the lowest life expectancy at birth and the population as whole by 2010.

Griffiths and Fitzpatrick acknowledged the limitations of presenting life expectancy results for local authorities without examining variation within authorities, some of which have very large populations e.g. Manchester and Birmingham. They noted that research at parliamentary constituency and ward level has revealed substantial variations in mortality within local authorities.[4,5] Other sources of data, such as the Index of Multiple Deprivation, show large differences in population characteristics within local authorities.

Demand for indicators of mortality for smaller areas, for monitoring purposes and to inform local strategies, have recently led some Public Health Observatories to calculate life expectancy for wards within their regions. For example, the East Midlands (formerly Trent) observatory has published maps of five counties with wards shaded to illustrate areas of the highest and lowest life expectancy.[6]

In 2001 ONS launched the Neighbourhood Statistics Service (NeSS), with the aim of supplying the information needs for the National Strategy for Neighbourhood Renewal. The Neighbourhood Renewal Unit in its identification of policy requirements for NeSS listed life expectancy at ward level as Priority A – essential. Concern was noted however regarding the potential methodological problems involved in its calculation.

In 2002 ONS therefore began a research project to consider the possibility of calculating life expectancy at birth for all wards in England and Wales, using a consistent methodology. This report summarises the options considered, the methods of testing that were employed, and the final methodological conclusions that were reached.

2. Research objectives

ONS has recently published a number of reports with results for life expectancy at birth for sub-national areas of the United Kingdom. [2,7,8,9] These results, published down to local authority level, were produced using three year aggregates of deaths and mid-year population estimates. Figures were calculated using life tables constructed by the established methodology developed by Chiang[10,11] and also illustrated in Newell,[12] and Shyrock and Siegel.[13]

While suitable for the calculation of life expectancy at local authority level it was not clear that the Chiang method would be the most suitable to use when considering calculation at ward level. In 2001 the mean population of local authorities in England and Wales was 139,000, compared to a mean ward population of around 5,800. In reports of life expectancy at birth at local authority level ONS has not published results with confidence intervals. For wards however some measure of variance was considered essential to allow for random variation in the annual number of deaths in areas with small populations (although deaths will be aggregated for a number of years).

A methodology was therefore required which would allow the calculation of a standard error for the life expectancy at birth result, which could be used to publish confidence limits. Even if research indicated that producing national ward level life expectancy results would not be a feasible option for ONS, it was considered that the establishment of a method for providing a measure of variance would still give added benefit to the other sub-national results that will continue to be published.

Although the mean population size of wards approached 6,000 in England and Wales in 2001, many wards had populations which were considerably smaller. It was therefore also an aim of the research study to establish if there was a population size which might be considered as a minimum, below which the calculation of results would not be considered viable.

It was expected that one particular problem with calculating life expectancy, and its variance, for small populations would be the effect of having no deaths in some age groups. It had been noted when calculating local authority level results that many areas had age groups where there were no deaths, especially at younger ages. At ward level the occurrence of zero deaths could be expected to be much more frequent.

Each age interval contributes to the standard error of life expectancy at birth. It was therefore suspected that having no deaths in an age band might have the effect of decreasing the measure of variance, because zero deaths would lead to a zero contribution to the variance from that particular interval. Research was also planned therefore to consider the potential underestimation of variance caused in these cases, and the possible means of countering this, such as by placing a small positive value in those age bands with no deaths.

The three key strands of the research study were therefore to:

1. Compare methodologies and select one suitable for use with small populations.
2. Establish a minimum population size below which life expectancy would not be calculated.
3. Consider the effect of having no deaths in age bands and test the impact of inserting other values.

3. Comparison of methodologies

ONS has published life expectancy results based on life tables constructed using Chiang's method of abridged life tables.[10,11] Abridged life tables aggregate deaths and population data into age groups, rather than using single year of age as in complete life tables. For local authorities ONS has calculated results based on the age groups under1, 1–4, 5–9…80–84, 85 and over. The same age bands were maintained in the analysis undertaken for this research study. Complete life tables are used by the Government Actuary's Department when calculating life expectancy at national level, but their use would not be appropriate at local authority or ward level.

Age-specific mortality rates are used within the life table to calculate the probability of dying at each age interval. These probabilities are then applied to a hypothetical population cohort of newborn babies. Life expectancy at birth in an area can therefore be defined as an estimate of the average number of years a new-born baby would survive if he or she experienced the particular area's age-specific mortality rates for that time period throughout his or her life.

An example of a life table constructed according to the Chiang methodology can be found in an Excel workbook on the National Statistics website at:

http://www.statistics.gov.uk/statbase/Product.asp?vlnk=8841

This file, 'Life Table Templates' contains annotated examples of the life tables compared in this research study.

The Chiang method is long established and has been widely used internationally. It is documented in a WHO publication and elsewhere.[10, 11,12,13] A more recent method was developed by Silcocks, Jenner and Reza and published in 2001.[14] The aim of Silcocks *et al* was to investigate the usefulness of life expectancy as a summary measure of mortality at health authority level or below, and to derive a formula for the variance of life expectancy results. As the Silcocks method was designed with the intention of using it to calculate life expectancy for sub-national populations, the research study was restricted to comparing this method with the established Chiang method.

An obstacle in calculating variance with the established Chiang method is the occurrence of age intervals where there are no deaths. This causes the calculation of the standard error to fail. However an alternative method of calculating the variance of the probability of death, and the consequent possibility of calculating the standard error using the Chiang method was identified.[15] A life table was also constructed using this methodology. There are therefore two Chiang life tables in the 'Life Table Templates' on the NS website. They are identical in their calculation of life expectancy but differ in their calculation of variance. The original Chiang methodology is labelled 'Chiang I', while the revised method is referred to as 'Chiang II'. A third life table example is included in the templates. This illustrates the Silcocks method and is based on a spreadsheet supplied to ONS by David Jenner.

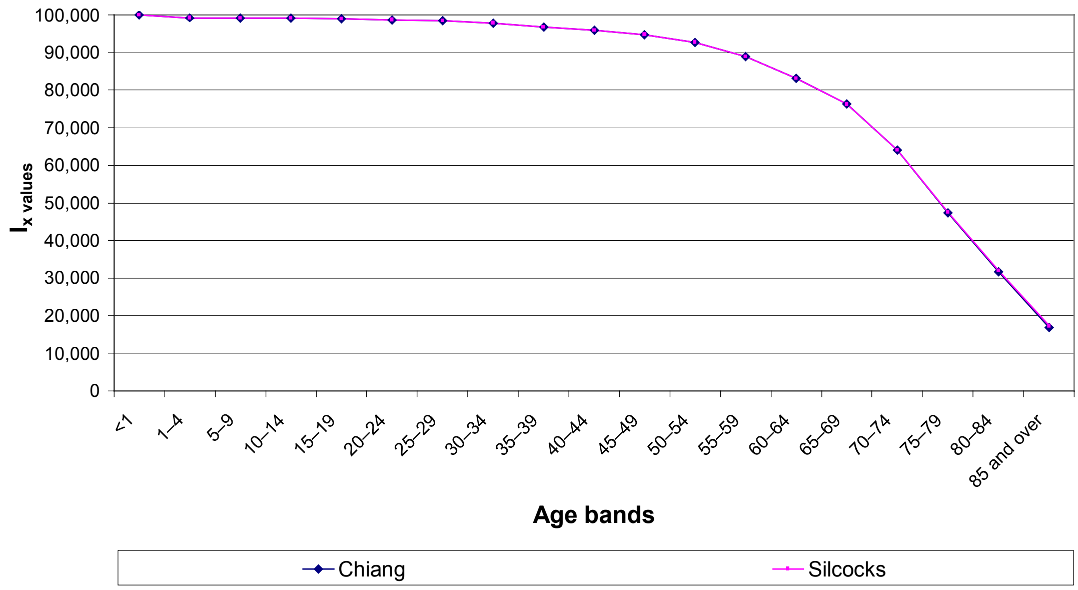
The Chiang methods use values which determine the fraction of the age interval lived by the population cohort who die in each age interval. In the life table templates these are the values labelled a_x. Chiang has published a_x values based on different populations at different time points to adapt the method to different countries. In the life tables used by ONS to calculate life expectancy for local authorities and larger geographic areas, a_x values of 0.5 are used for every age interval expect those under 1. This therefore assumes that for all those dying in an age interval the deaths were spread evenly throughout the period. The average number of years lived within each interval by those who died can therefore be assumed to be half the width of the age band.

The same distribution cannot be assumed for infant deaths however, the majority of which occur in the first 28 days of life. ONS therefore uses the convention of assigning deaths of under 1s an a_x value of 0.1 These a_x values were maintained in the life tables used in the research process. The Silcocks method does not explicitly use a_x values in the life table but it assumes a_x is 0.5 for all age groups in the formulae for calculating life expectancy and its variance.

While the Chiang methods assume that deaths are spread evenly throughout each age period the Silcocks method assumes that the mortality rate in each age group is constant. The probability of survival therefore decreases exponentially through the age band.

Life expectancy at birth calculated using the Chiang methods and the Silcocks method were compared and found to be very similar. For each a hypothetical cohort population of 100,000 was assumed. The number of individuals out of these original 100,000 who were still alive at each age band were compared. These are the values labelled l_x in the life table templates. Figure 1 plots an example of these l_x values and illustrates that the Chiang and Silcocks values are almost identical. The very small differences that do arise are the result of slightly different l_x calculations due to the Silcocks method of exponential approximation. Life expectancy at birth is simply the sum of the area under the curve, and so the results in this example are almost indistinguishable.

Figure 1 Chiang and Silcocks methods – comparison of l_x values



The Chiang and Silcocks methods differ also in the measurement of variance for the final age interval. Chiang assumes zero variance for the final age band as there can be no probability of survival. The Silcocks method however does include an estimate of variance which takes the final age interval into account based on length of survival rather than probability of survival.

To consider the effect of taking the final age interval into account, in the life table templates available on the website, an adjustment has been made to the Chiang life tables. These illustrate the standard error calculated with the assumption of zero variance for those aged 85 and over, and also an adjusted figure calculated using the suggested Silcocks method for the final age band. The life tables illustrate that the adjustment does not affect the calculation of life expectancy at birth, only its associated standard error.

Given the similarity in life expectancy at birth calculated using the Chiang And Silcocks methods, and as the former is already being used by ONS to calculate results at local authority level, it was decided that this method would be adopted for the potential calculation of life expectancy at ward level. The further two aims of the research process, to consider a minimum population size and the effect of zero deaths in age bands, were therefore conducted using just the Chiang II method. This allowed the calculation of variance in the presence of age bands with zero deaths.

4. Establishing a minimum population size

In order to attempt to establish a minimum population size, below which life expectancy calculations would not be considered viable, a range of populations were tested. These are listed in Box A.

Box A

Population size
1,000
2,000
3,000
4,000
5,000
7,500
10,000
20,000
30,000
40,000
55,000
80,000

Table 1 also shows the average population sizes of wards and local authorities in England and Wales, based on results from the 2001 Census.

The number of wards does not include the 18 in England and Wales for which individual population counts from the 2001 Census have not been published because their population threshold was too low. 16 of these were in the City of London with the other two in the Isles of Scilly and Lancaster. ONS has not published life expectancy at birth results for the City of London and Isles of Scilly local authorities on the basis that their small populations and numbers of deaths would not produce sufficiently robust results. These two areas have also been excluded from the local authority figures in Table 1. This table demonstrates how much more challenging it will be to calculate meaningful life expectancy results for wards, as compared to local authorities.

Although the mean ward population size approaches 6,000 it can be seen that at least 10 per cent of wards have a sex-specific population of less than 1,000. This was still chosen as the minimum population size to test, to allow for the fact that data are likely to be aggregated for a number of years. The population size was increased by 1,000 for each series of tests, until 5,000, when the intervals were then increased.

Table 1 Ward and local authority population distribution, England and Wales, 2001

	Number of areas	Mean population	Median population	Population below which 10% of geographic areas lie	Population above which 10% of geographic areas lie
Wards					
Persons	8,850	5,880	4,820	1,910	11,890
Males	8,850	2,860	2,350	930	5,770
Females	8,850	3,020	2,480	970	6,090
Local authorities					
Persons	374	139,240	113,010	66,710	246,000
Males	374	67,780	55,740	32,300	120,240
Females	374	71,460	58,250	34,440	125,490

Methods

The tests were conducted using Monte Carlo simulations, which were used to generate life expectancy figures from a fixed population using a sample distribution. The aim of running the simulations was to provide evidence of how life expectancy at birth, and its standard error, are distributed for a variety of population sizes so that, for example, there would be an indication of how wide expected confidence intervals might be for different sized wards.

The Chiang II method was used to calculate life expectancy and its standard error and no adjustment was made to assess the contribution of variance in the last age interval. Underlying death rates were based on real death rates that were given in English Life Tables (ELTs). Every ten years the Government Actuary's Department produces a set of graduated (smoothed) life tables for England and Wales, known as the English Life Tables. Each of the English Life Tables is associated with decennial population censuses, beginning with the census of 1841. ELT 15 is the 15th in the series and contains graduated life tables constructed from the mortality experience of the population of England and Wales during the three years 1990, 1991 and 1992.

Underlying death rates were based on English Life Tables 11 to 15, in order to represent a range of life expectancy scenarios from lower to higher (reflecting the wide disparity already identified between life expectancy results for local authorities). Deaths were randomly generated (using a binomial distribution) for the different population sizes listed in Box A, with the structure of the population fixed according to the 1991 mid-year population estimates for England and Wales. The underlying death rates were therefore used to calculate plausible results for a range of life expectancy experiences. These provided 'reference' values against which simulated results could be compared.

Computer programs using Stata[16] were run to generate Monte Carlo simulations. For each combination of age-specific death rates and population size the simulations were repeated 2,000 times, for both males and females. The mean of the life expectancy results was calculated for each combination and compared to the reference life expectancies calculated using the mortality rates from the English Life Tables. The same was done for the standard errors.

Simulation results

The mean of the life expectancy at birth results from the simulations for a selection of the population sizes tested are presented in Table 2. They are compared with the reference life expectancy results. The reference results do not alter by population size as they are based on the underlying death rates from the English Life Tables. The simulations were run for males and females but only results for the latter are shown in this and the following tables.

The results from the simulated data for all population sizes show that life expectancy approximated well to the reference results for each English Life Table considered, although the reference results are slightly underestimated in the majority cases. However for the smallest populations tested the results from the simulations were more likely to be overestimated when compared to the reference results. Comparable results were observed for males.

Table 2 Female life expectancy at birth – reference results and simulation comparisons.

Years

English Life Table	Reference life expectancy	Life expectancy from simulations by population size								
		80,000	40,000	20,000	10,000	5,000	4,000	3,000	2,000	1,000
ELT 11	71.93	71.89	71.87	71.80	71.87	71.78	71.63	71.94	72.19	72.43
ELT 12	74.27	74.23	74.23	74.22	74.20	73.97	74.24	74.39	74.60	74.67
ELT 13	75.40	75.40	75.34	75.43	75.33	75.38	75.34	75.40	75.52	75.87
ELT 14	76.91	76.89	76.89	76.91	76.77	76.86	77.01	76.83	77.10	77.66
ELT 15	78.24	78.21	78.21	78.13	78.19	78.30	78.28	78.29	78.21	78.41
		Differences in life expectancy between reference and simulation results								
		80,000	**40,000**	**20,000**	**10,000**	**5,000**	**4,000**	**3,000**	**2,000**	**1,000**
		-0.04	-0.06	-0.13	-0.06	-0.14	-0.29	0.02	0.26	0.51
		-0.04	-0.04	-0.05	-0.07	-0.30	-0.03	0.12	0.33	0.40
		0.00	-0.06	0.03	-0.07	-0.02	-0.05	0.00	0.12	0.47
		-0.01	-0.02	0.01	-0.14	-0.05	0.10	-0.07	0.19	0.76
		-0.02	-0.02	-0.10	-0.05	0.06	0.05	0.06	-0.03	0.17

The standard errors for the life expectancy results in Table 2 are presented in Table 3. Unlike the life expectancy reference results, which are not affected by changing the size of the population, the standard errors are strongly related to population size. Table 3 therefore includes the standard errors calculated for each population size and each set of underlying death rates based on English Life Tables 11 to 15. These are compared with the standard errors generated from the simulations. The standard errors from the simulations approximate well to the reference data standard errors for larger population sizes. As the population size decreases though, especially under 5,000, the standard error tends to be underestimated.

Table 3 Standard errors for female life expectancy at birth – reference results and simulation comparisons

Years

English Life Table	Standard errors from reference data life expectancy, by population size								
	80,000	40,000	20,000	10,000	5,000	4,000	3,000	2,000	1,000
ELT 11	0.57	0.81	1.14	1.62	2.28	2.56	2.95	3.62	5.11
ELT 12	0.53	0.75	1.06	1.50	2.11	2.37	2.73	3.35	4.73
ELT 13	0.51	0.72	1.03	1.45	2.05	2.30	2.65	3.25	4.59
ELT 14	0.48	0.67	0.95	1.35	1.90	2.13	2.46	3.02	4.26
ELT 15	0.44	0.62	0.88	1.25	1.76	1.97	2.28	2.79	3.94
English Life Table	**Standard errors from simulated data life expectancy, by population size**								
	80,000	**40,000**	**20,000**	**10,000**	**5,000**	**4,000**	**3,000**	**2,000**	**1,000**
ELT 11	0.57	0.81	1.15	1.61	2.27	2.56	2.85	3.40	4.57
ELT 12	0.53	0.75	1.05	1.49	2.12	2.31	2.61	3.11	4.21
ELT 13	0.51	0.73	1.02	1.45	2.01	2.25	2.56	3.08	4.06
ELT 14	0.48	0.67	0.95	1.36	1.87	2.07	2.38	2.85	3.73
ELT 15	0.44	0.62	0.89	1.24	1.72	1.90	2.19	2.66	3.53
	Differences in standard errors between reference and simulation results								
	80,000	**40,000**	**20,000**	**10,000**	**5,000**	**4,000**	**3,000**	**2,000**	**1,000**
	0.00	0.00	0.00	-0.01	-0.01	0.00	-0.10	-0.22	-0.54
	0.00	0.00	0.00	-0.01	0.01	-0.06	-0.12	-0.24	-0.52
	0.00	0.00	0.00	0.00	-0.04	-0.05	-0.09	-0.17	-0.53
	0.00	0.00	0.00	0.01	-0.04	-0.06	-0.08	-0.17	-0.53
	0.00	0.00	0.01	-0.01	-0.04	-0.07	-0.09	-0.13	-0.41

In summary, Tables 2 and 3 indicate that compared to the reference results for the smallest populations, the simulation results for life expectancy were likely to be overestimated, while standard errors were likely to be underestimated.

The cause of each effect is probably related to the progressive increase in the number of age intervals with zero deaths, as population size decreases.

The standard error results from the reference data presented in Table 3 are also illustrated in Figure 2, where the relationship between population size and standard error is clearly evident. As the population size decreases the standard error increases and consequently the width of confidence intervals calculated with these would also increase.

The width of these confidence intervals are illustrated in Table 4. 95% confidence limits were calculated (1.96 x standard error) for female life expectancy at birth reference results based on the underlying deaths rates from English Life Table 15.

Figure 2 Standard errors for female life expectancy at birth – reference results

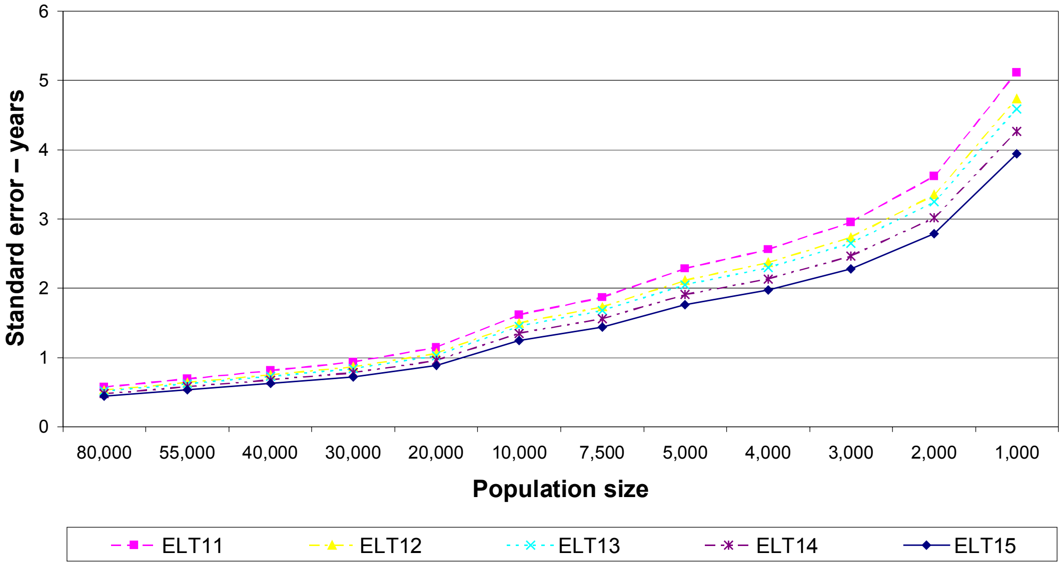


Table 4 Female life expectancy at birth – reference results based on English Life Table 15, with 95% confidence intervals, by population size

Life expectancy at birth reference result = 78.24 years

Years

	Population								
	80,000	**40,000**	**20,000**	**10,000**	**5,000**	**4,000**	**3,000**	**2,000**	**1,000**
Lower confidence limit	77.37	77.01	76.51	75.79	74.78	74.37	73.77	72.77	70.51
Upper confidence limit	79.10	79.46	79.96	80.68	81.69	82.10	82.70	83.70	85.96
Width of confidence interval	1.73	2.44	3.46	4.89	6.90	7.73	8.93	10.93	15.44

The width of the confidence interval (width between the upper and lower confidence limits) increases as the population decreases. The widths indicated that life expectancy at birth calculated for populations of less than 5,000 were likely to be in excess of seven years. This was an important factor when considering a minimum population size. For the population of 1,000 the interval width was over 15 years. If the purpose of publishing confidence intervals is to allow the identification of areas which are significantly higher or lower than others, then interval widths of this size would be very unlikely to permit this.

5. Age bands with zero deaths

In their paper Silcocks *et al* [14] noted that in populations with no deaths in some age bands, there will be a 'spurious decrease in the estimated sampling variation' – the standard error will be underestimated because the contribution to the variance for those intervals with no deaths is equal to zero. This is a problem that arises even when calculating life expectancy at local authority level, and will be greatly exaggerated in any attempt to produce ward level life expectancies. Silcocks *et al* explored the problem of zero deaths by correcting for age bands with "random zeroes" by imputing a small positive value. They tested these using a simulated population of 28,000. The simulations assumed a Poisson distribution of deaths within each age interval. Two alternative imputed values were tested for age groups with zero deaths:

- 0.693 – the value of the Poisson rate for which the chance of observing 0 events is a half.
- 3 – the upper 95% confidence limit for the frequency of events when none are actually observed.

The corrections were found to increase the standard error. However only a handful of wards in England and Wales had populations over 28,000 in 2001. With the mean population of wards around 5,800 persons, there was concern that imputing small values into very small populations would increase the standard error but also result in large decreases in life expectancy.

To extend the work of Silcocks *et al* a range of possible values were considered using a similar approach to that documented in their paper. Monte Carlo simulations, which had been used to assess the distribution of life expectancy and its standard error for different population sizes, were run again but with constant values inserted in place of zero where there were no deaths in an age band.

The simulations were first run by imputing the smaller of the two values tested by Silcocks *et al* (0.693) into every age interval with no deaths. The resulting comparisons with results generated from English Life Table 15 are presented in Table 5 for life expectancy, and Table 6 for standard errors. For populations of 10,000 and smaller, imputing a value of 0.693 reduced the resulting life expectancy by over two years. This suggested that this value would be too high to use even for the larger wards.

Table 5 Female life expectancy at birth, results from data simulations based on English Life Table 15, with 0.693 imputed into age bands with zero deaths

Years

	Population size								
	80,000	40,000	20,000	10,000	5,000	4,000	3,000	2,000	1,000
Reference life expectancy	78.24	78.24	78.24	78.24	78.24	78.24	78.24	78.24	78.24
Simulated life expectancy	78.14	77.91	77.26	75.89	72.86	71.26	68.67	63.71	51.13
Difference (years)	-0.10	-0.32	-0.97	-2.35	-5.38	-6.98	-9.57	-14.53	-27.10
Percentage Difference	-0.13	-0.41	-1.24	-3.00	-6.87	-8.92	-12.23	-18.57	-34.64

Table 6 Standard errors for female life expectancy at birth, results from reference data and from simulations based on English Life Table 15, with 0.693 imputed into age bands with zero deaths

Years

	Population size								
	80,000	40,000	20,000	10,000	5,000	4,000	3,000	2,000	1,000
Standard error from reference data life expectancy	0.44	0.62	0.88	1.25	1.76	1.97	2.28	2.79	3.94
Standard errors from simulated life expectancy with imputation	0.45	0.65	0.99	1.59	2.66	3.16	3.93	5.25	7.81
Difference (years)	0.01	0.03	0.11	0.34	0.90	1.19	1.65	2.46	3.87
Percentage difference	1.63	4.56	12.70	27.35	51.30	60.38	72.58	88.12	98.36

As 0.693 appeared too large a value to use for small populations, a range of smaller positive values were also tested. These are listed in Box B. Also tested was the imputation of a value based on the sex-specific rate for England and Wales for each age group. Simulations were recreated using all the population sizes previously tested (1,000–80,000) and with the underlying mortality rates again based on English Life Tables 11 to 15.

Figure 3 shows the life expectancy at birth results from the simulations for each imputed value, for a selection of population sizes. As might be expected, the larger the imputed value and the smaller the population size, the further the distance from the simulated life expectancy to the result from the reference data. For all population sizes, the smaller the imputed value, the closer the resulting life expectancy to the reference figure. Imputing a value based on national sex-specific mortality rates produced very similar life expectancy and standard error results to those when no values were imputed.

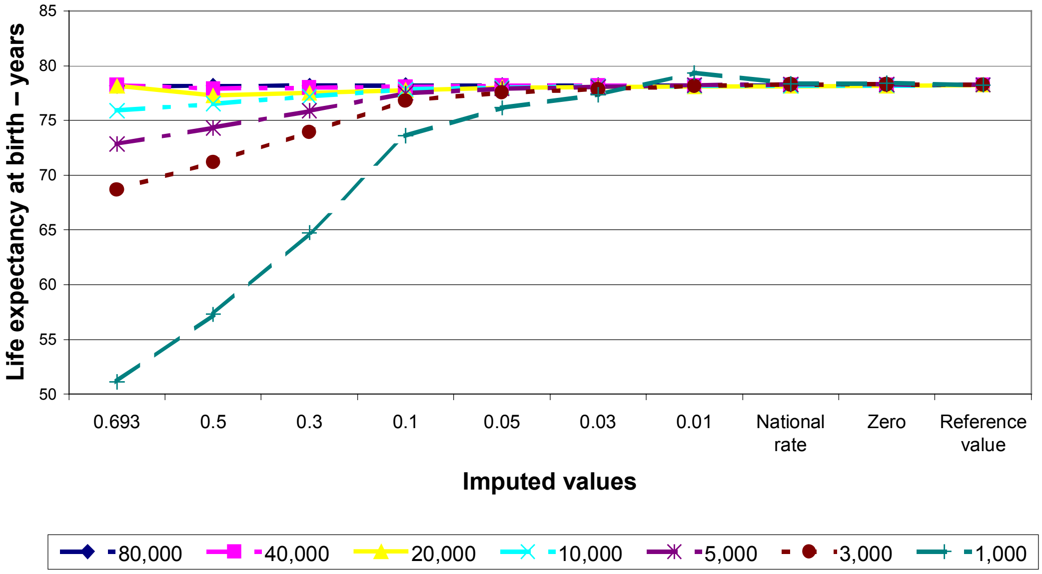
Box B

Values for imputation testing

0.010
0.030
0.050
0.100
0.300
0.500
0.693

national sex–specific mortality rate

Figure 3 Female life expectancy at birth, simulated data based on English Life Table 15, by different imputed values and population sizes



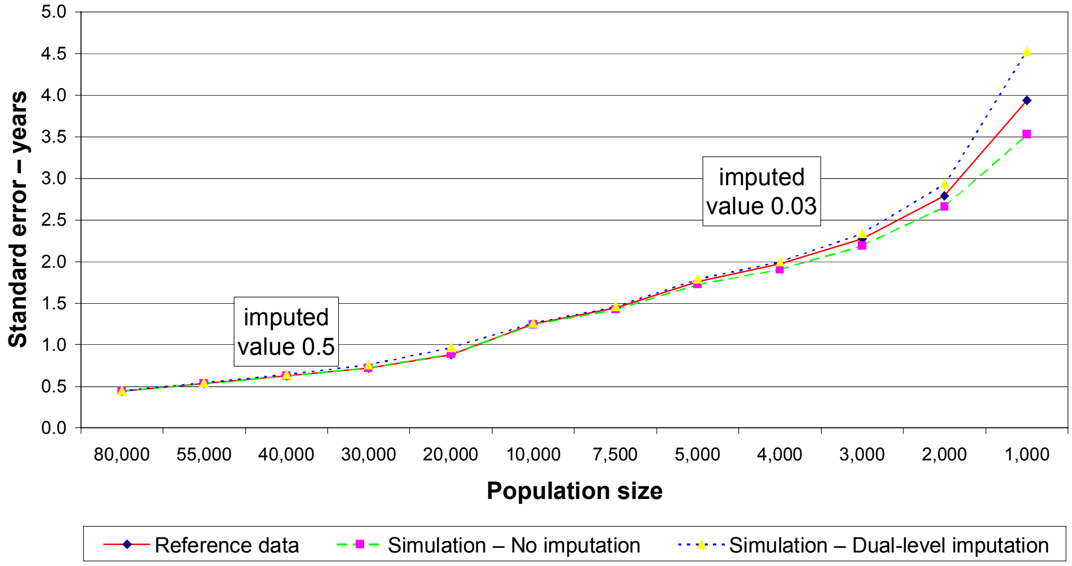
The results from the simulations, which tested a range of imputed positive values, suggested that while larger values could be imputed for larger populations, the same values could not be used for smaller populations without greatly reducing their life expectancy. A different scale was therefore considered which varied the imputed value according to the size of population. In the example illustrated in Figure 4, 0.5 was used for populations of 20,000 upwards and 0.03 for the smaller population sizes. These two values gave a good approximation to the reference life expectancy for the range of populations considered.

Figure 4 presents standard error figures calculated for female life expectancy at birth based on underlying deaths rates from English Life Table 15. The middle line is the standard errors obtained from the reference data. The top line is the standard errors obtained from data simulations when 'dual level' imputation was used – imputing 0.03 into age bands with zero deaths for populations less than 20,000 and imputing 0.5 into age bands for populations of 20,000 and over. The simulations were also run again without imputing any value into age intervals where there were no deaths. These results are shown in the bottom line.

Results are presented just for female life expectancy based on English Life Table 15. Comparable figures were also obtained however from data simulations based on English Life Tables 11 to 14, and for males.

Figure 4 illustrates that for populations of less than 5,000 imputing even a very small value such as 0.03 will tend to overestimate the standard error. Conversely, leaving age bands with zero deaths will lead to an underestimation of the standard error for populations smaller than 5,000. For larger populations however, from 5,000 upwards, the results without imputation approximate closely to the reference data. This indicates that the presence of age bands with no deaths is likely to have little impact on the calculation of standard errors for populations of 5,000 and over.

Figure 4 Standard errors for female life expectancy at birth based on English Life Table 15



These results suggest that for all but the smallest populations a good approximation to the standard error produced from the reference data is likely without imputing any value into age bands with no deaths. For the final age interval however the presence of zero deaths causes the Chiang II method of calculating variance to fail.

Imputing a value based on the national age and sex specific death rate was therefore adopted in cases where there were no deaths in the final age interval as it had been observed that this produced results which were very similar to those when no values were imputed.

6. Adjusting for the final age interval

The data simulations which were run to inform decisions regarding a minimum population size and the effect of age bands with no deaths, were calculated without adjusting for variance in the final age band. Additional analysis was therefore needed to determine whether or not an adequate approximation of variance could be made without adjustment for the final age band, or whether an estimation of variance which considers the final age interval should be made, as suggested by Silcocks *et al*.

The adjustment to the Chiang methodology, with the Silcocks method for taking account of the last age group, is illustrated in the life table templates on the National Statistics website (see page 7). This adjustment was used in data simulations run using the methods described earlier in Section 4. For this analysis the simulations were restricted to underlying death rates based just on English Life Table 15. Estimates of standard errors were calculated using the Chiang II method both with and without adjustment for the final age interval.

The results from the simulations approximated very well to the reference figures which are presented in Table 7. Even with a population of 1,000 the adjustment only increased the standard error estimate by 0.05 years. For populations of 5,000 and over the adjustment increased the standard error by 0.02 years or less.

Table 7 Standard errors for female life expectancy at birth – reference results based on English Life Table 15

Years

	Population size								
	80,000	**40,000**	**20,000**	**10,000**	**5,000**	**4,000**	**3,000**	**2,000**	**1,000**
Standard error **without** variance for the last age interval	0.44	0.62	0.88	1.25	1.76	1.97	2.28	2.79	3.94
Standard error **with** variance for the last age interval	0.45	0.63	0.89	1.26	1.78	2.00	2.31	2.83	3.99
Difference	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.05

These results were based on a final age band of 85 and over. In their paper Silcocks *et al* caution against the use of ‘open-ended’ age groups as population structures can differ greatly. They advise using a fixed upper limit beyond which few deaths occur, such as age 95, or adopting an open-ended age group with a lower limit as high as possible. ONS has based its sub-national life expectancy results on age bands where the final interval is 85 and over. For larger geographic areas it may be appropriate for this upper age band to be raised, to 90 and over for example, or to adopt a fixed upper limit.

At ward level however any gains in a more sensitive calculation of age-specific rates for the over 85s would have to be balanced against the reliability of small area population figures for the very elderly. There would be an increased risk of introducing bias if the upper age interval was raised as the very elderly are likely to be particularly subject to migration effects, such as those associated with moves to care homes.

An alternative approach to data for the very elderly is to calculate area death rates only up to a fixed age limit, for example 85 or 90, and to estimate death rates over that limit using national death rates for the ‘open ended’ age group. There are a number of techniques available for making these estimates.

One technique is to ignore area variations and simply use the national estimates in place of area specific figures. Another is to model the relationship between age and mortality nationally and in each area at older ages. The national estimate of mortality in the open ended age group can then be adjusted to take account of the ‘modelled’ estimate of area specific mortality. For example, the ratio of modelled area mortality rate to the modelled national rate at ages 80 to 84 could be multiplied by the national mortality rate for ages 85 and over to estimate the area mortality rate at ages over 85.

Research has shown that the probability of dying generally increases rapidly with age until age 80.[17] At very old ages (90–100), the increase with age tails off, and appears to reach a plateau at around 105–110.[18,19] The mortality rates in all areas are therefore likely to tend to an ultimate level at extremely old ages (somewhere below a probability of dying in any year of 0.6).[19] For small areas, numerator-denominator biases and random fluctuations in deaths in the oldest age group are therefore likely to exceed any underlying variation in mortality rates between these areas at the oldest open ended age range. Applying an unadjusted national rate to all wards would reduce the risk of bias in areas particularly affected by migration effects in very elderly populations. Even an adjusted figure would have some effect in reducing this bias.

7. Methodological Conclusions

The results of the data simulations provided an informed basis on which decisions regarding the methodological options could be made. The revised Chiang methodology (Chiang II), used to calculate life expectancy at birth with standard errors in all the simulations, proved to be a robust method which ONS will use for the calculation of life expectancy at all geographic levels.

Results from Tables 2 and 3 indicated that for very small populations, under 5,000, life expectancy at birth results tended to be overestimated, while standard errors for these results tended to be underestimated. When an adjustment was made for the underestimation of variance by imputing a small positive value into age bands with no death, standard errors were then inclined to be overestimated in populations of less than 5,000 (Figure 4).

The width of the 95% confidence intervals for different population sizes illustrated in Table 4 also indicated that for populations of 4,000 they were likely to be approaching eight years and for populations of 1,000 the likely interval width would be over 15 years.

Based on these results it is clear that consistently reliable and robust calculations of life expectancy at birth, and a measure of its variance, cannot be made for populations of less than 5,000. It has therefore been accepted that 5,000 will be the minimum population base below which life expectancy calculations will not be attempted by ONS. The data simulations revealed that for populations of 5,000 and over, life expectancy and standard error results approximated well to the reference data based on real underlying mortality rates from the English Life Tables.

The simulations also indicated that for a population of 5,000 the width of 95% confidence intervals was still likely to be around seven years. It is likely that this will still allow the identification of wards which have significantly different life expectancies. At local authority level the difference in male life expectancy at birth between areas with the highest and lowest results is now around ten years. Differences at ward level may be even greater. In his annual report in 2001[20] the Chief Medical Officer of the Department of Health identified that some wards in England in 1996–1998 had mortality rates which matched national average death rates from the middle of the Twentieth Century.

The data simulations also provided evidence for an approach to take to populations which have age groups where no deaths occurred. The results suggested that using the revised Chiang method, the closest approximations to reference life expectancy and standard errors would be achieved by not imputing any values into age groups with zero deaths. For the final age band a value is essential however, otherwise the calculation of variance fails. In these cases a value based on the sex specific rate for those aged 85 and over in England and Wales will be inserted.

For the forthcoming ward level life expectancy results that ONS intends to release through the Neighbourhood Statistics Service (NeSS) the upper age interval will be maintained at 85 and over. ONS will however consider further methodological approaches to estimating mortality in the upper age band for these calculations, including applying a national rate, above a fixed age limit, for all wards. Extending the upper age interval, or using a fixed upper age limit, will also be considered for local authorities and larger geographies. ONS will assess the impact this would have before any change is made to the current methodology.

The methodological options concluded by ONS can thus be summarised:

1. Calculation of sub-national life expectancy at birth in the United Kingdom, and a measure of its variance, will be based on Chiang's revised methodology.

2. ONS will not attempt to calculate life expectancy at birth for a population below 5,000.

3. The occurrence of age bands where no deaths occur will not be adjusted for by imputing any other value. The final age band will however be adjusted in these cases by inserting a value based on national sex-specific death rates.

4. The upper age band for the calculation of ward level life expectancy will be 85 and over for forthcoming results to be released through NeSS. ONS will consider further whether this is also the most appropriate upper age band for the calculation of life expectancy for larger populations, and whether alternative approaches can also be applied at ward level.

Life tables illustrating the methodologies tested in this research study are available in an Excel workbook, 'Life Table Templates' on the National Statistics website at:

http://www.statistics.gov.uk/statbase/Product.asp?vlnk=8841

The life table 'ONS Methodology' contains the method chosen by ONS to calculate life expectancy at birth at all geographic levels.

8. Application of the methodology

This research study has established a methodology which will allow ONS to calculate life expectancy at birth, with 95% confidence intervals, for populations of 5,000 and more. This method has now been used to add confidence intervals to previously published life expectancy results on the National Statistics website at:

http://www.statistics.gov.uk/statbase/Product.asp?vlnk=8841

There remain some challenges however to the publication of ward level life expectancy results in England and Wales.

Table 1 illustrated the population distribution among wards in England and Wales in 2001. Although the mean population approached 6,000 the mean sex-specific populations were around 3,000. In addition 10 per cent of wards had sex-specific populations which were less than 1,000. To reach the minimum population threshold of 5,000 it will therefore be necessary to aggregate populations. There are three principal options:

1. Aggregation of sex-specific populations – It is clear that calculating life expectancy results for persons rather than males and females would greatly increase the numbers of areas which would meet the minimum population. A person based life expectancy might also be considered a reasonable indication of the mortality experience of the entire population of a small area. However at national level there is a gap of around 5 years between life expectancy at birth between males and females. Figures for local authorities have also indicated that figures for both sexes are not always comparable, e.g. an area with very low female life expectancy may not necessarily have very low male life expectancy. In order not to loose this level of detail ONS intends to calculate sex-specific life expectancies and to consider other aggregation strategies which will allow it to do so.

2. Temporal aggregation – Recent life expectancy at birth results for local authorities have all been based on three year aggregations of population estimates and mortality data. Results for wards would also be based on at least three years of data and it may be necessary to consider aggregating over five years. Using the most recent five years of mortality data would base results on deaths registered between 1998 and 2002. Data for 1998 would therefore be three years away from the 2001 Census populations. This introduces a risk of increasing possible numerator/denominator discrepancies, particularly for very small populations.

3. Geographic aggregation – Aggregating data even over a five year period would still leave some wards in England and Wales with sex-specific populations of less than 5,000. It would therefore be necessary to consider aggregating some wards geographically. Rather than combining areas based purely on their geographic proximity ONS will consider methods which would allow the aggregation of areas based on shared characteristics. These may include considering wards within local authorities which share similar mortality experiences based on the calculation of Standardised Mortality Ratios. The ONS ward area classification, which classifies areas according to shared socio-economic and demographic characteristics may also be considered as an aggregation tool. The ward area classification based on the results of the 2001 Census is likely to be released by ONS towards the end of 2003.

As some wards have a sex-specific population of less than 1,000 (as shown in Table 1) even if data were aggregated over five years there would still be a need for some geographic aggregation. The number of areas which would need to be geographically aggregated would increase if data were aggregated over only three years. ONS intends to publish a life expectancy at birth figure for every ward in England and Wales (apart from wards in the City of London and Isles of Scilly). For some wards however their results will be based on data which has been aggregated with data from one or more other wards. These cases will be indicated in published results.

ONS plans to provide ward level life expectancy at birth results for England and Wales through the Neighbourhood Statistics Service. Although ONS has published life expectancy figures for local authority areas in Scotland and Northern Ireland in recent reports (and is likely to continue to produce these in the future) it does not intend to produce ward level life expectancy results for these countries.

In some cases local biases may influence ward level life expectancy results. For example, the presence of nursing homes in a ward may lead to local migration effects which will influence mortality rates. The interpretation of ward level life expectancies will in some cases therefore require local knowledge. ONS will aim to facilitate interpretation by flagging wards which have a high proportion of their population resident in nursing homes, using data from the 2001 Census.

References

1. *Fifth Annual Report of the Registrar General*, 1843. HMSO: London.
2. Griffiths C and Fitzpatrick J (2001) Geographic inequalities in life expectancy in the United Kingdom, 1995–1997. *Health Statistics Quarterly* **09**, 16–28.
3. Tackling Health Inequalities – Consultation on a plan for delivery (2001) Department of Health.
4. Shaw M, Dorling D, Gordon D and Davey Smith G (1999) The widening gap: health inequalities and policy in Britain, The Policy Press: Bristol.
5. William E S, Scott C M and Scott S M (1995) Using mortality data to describe geographic variations in health status at sub-district level. *Public Health* **109**, 65–73.
6. East Midlands Public Health Observatory. Illustrations of health inequalities in the East Midlands: http://www.empho.org.uk/products/inequalities/reports.htm
7. Office for National Statistics (2001) Life expectancy at birth by health and local authorities in the United Kingdom, 1997–99. *Health Statistics Quarterly* **11**, 78–85.
8. Office for National Statistics (2002) Life expectancy at birth by health and local authorities in the United Kingdom, 1998 to 2000. *Health Statistics Quarterly* **13**, 83–90.
9. Life expectancy at birth by health and local authority in England and Wales, 1991-1993 to 1999-2001 – revised figures using 2001 Census based population estimates (2003): http://nswebcopy/statbase/Product.asp?vlnk=8841
10. Chiang C L (1978) Life Table and Mortality Analysis, World Health Organisation.
11. Chiang C L (1984) The Life Table and its Applications, Malabar (FL), Robert E Krieger Publ Co.
12. Newell C (1994) Methods and Models in Demography. John Wiley & Sons: Chichester.
13. Shyrock H S and Siegel J S (1976) The Methods and Materials of Demography. Abridged edition, Academic Press: New York.
14. Silcocks P B S, Jenner D A, Reza R (2001) Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities, *Journal of Epidemiology and Community Health* **55**, 38–43.
15. Chiang C L, The life table and its construction. From: Introduction to stochastic processes in Biostatistics. New York: John Wiley & Sons 1968, Chapter 9, 189–214.
16. Stata statistical software (Release 7) – StataCorp, College Station, TX, USA.
17. Thatcher AR. (1999) The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society* (A) **162**, 5–43.
18. Thatcher A R (1999) The demography of centenarians in England and Wales. *Population Trends* **96**, 5–12.
19. Robine J-M and Vaupel J W (2001) Supercentenarians: slowly ageing individuals or senile elderly? *Experimental Gerontology* **36**, 915–930.
20. On the State of the Public Health – The Annual Report of the Chief Medical Officer of The Department of Health 2001. Department of Health.

Reports in the National Statistics Methodological Series

Report 1	Dave Elliot. *Software to weight and gross survey data with applications to the EC Household Panel and Family Expenditure Surveys*, (Government Statistical Service: 1997)
Report 2	*Report of the task force on seasonal adjustment*, (Government Statistical Service: 1996)
Report 3	*Report of the task force on imputation*, (Government Statistical Service: 1996)
Report 4	*Report of the task force on disclosure*, (Government Statistical Service: 1995)
Report 5	Peter Sharp. *GDP: Output methodological guide*, (Government Statistical Service: 1998).
Report 6	Michael A Baxter. *Interpolating annual data into monthly or quarterly data*, (Government Statistical Service: 1998)
Report 7	Dave Elliot. *Sample Design Options for an Integrated Household survey*, (Government Statistical Service: 1998)
Report 8	Kate Foster. *Evaluating non-response on household surveys*, (Government Statistical Service: 1998)
Report 9	Andrew Machin. *Reducing Statistical Burdens on Business*, (Government Statistical Service: 1998)
Report 10	David Ruffles. *Statistics on Trade in Goods*, (Government Statistical Service: 1998)
Report 11	Patrick Sturgis and Peter Lynn. *The 1997 UK Pilot of the Eurostat Time Use Survey*, (Government Statistical Service: 1998)
Report 12	Jeff Golland, David Savage, Tim Pike and Stephen Knight. *Monthly Statistics on Public Sector Finances – A Methodological Guide*, (Government Statistical Service: 1999)
Report 13	Gad Nathan. *A Review of Sample Attrition and Representativeness in Three Longitudinal Surveys,* (Government Statistical Service: 1999)
Report 14	Vera Ruddock. *Measuring and Improving Data Quality*, (Government Statistical Service: 1999)
Report 15	Peter Sharp. *Gross Domestic Product: Output approach (Gross Value Added)*, (Government Statistical Service: 1999).
Report 16	Dave Elliot. *Report of the Task Force on Weighting and Estimation*, (Government Statistical Service: 1999)
Report 17	P S Clarke and P F Tate. *Methodological Issues in the Production and Analysis of Longitudinal Data from the Labour Force Survey*, (Government Statistical Service: 1999)
Report 18	Margaret Frosztega and the Households Below Average Income team. *Comparisons of income data between the Family Expenditure Survey and the Family Resources Survey*, (Government Statistical Service: 2000)
Report 19	C Cousins, A Croft and E Snowden. *European Community Household Panel: Robustness Assessment Report for United Kingdom Income Data, Waves 1 and 2*, (Government Statistical Service: 2000)
Report 20	Ian Richardson. *Producer Price Indices: Principles and Procedures*, (Government Statistical Service: 2000)
Report 21	D J Holmes and C J Skinner. *Variance Estimation for Labour Force Survey Estimates of Level and Change*. (Government Statistical Service: 2000)

Reports in the National Statistics Methodological Series continued

Report 22	Michael White, Joan Payne and Jane Lakey. *Longitudinal Data for Policy Analysis*. (Government Statistical Service: 2000)
Report 23	Office for National Statistics Survey Control Unit. *Report on GSS Survey Activity in 1997*. (Office for National Statistics: 2000)
Report 24	Eileen Goddard. *Obtaining information about drinking through surveys of the general population*. (Office for National Statistics: 2001)
Report 25	Leicester Gill. *Methods for Automatic Record Matching and Linkage and their Use in National Statistics*. (Office for National Statistics: 2001)
Report 26	Sharon Bruce, Charles Lound and Dave Elliot. *Designing Surveys using variances calculated from Census data*. (Office for National Statistics: 2001)
Report 27	Office for National Statistics Survey Control Unit. Report on GSS Survey Activity in 1998. *(Office for National Statistics: 2001)*
Report 28	Ray Chambers. *Evaluation Criteria for Statistical Editing and Imputation.* (Office for National Statistics: 2001)
Report 29	Office for National Statistics Survey Control Unit. *Report on GSS Survey Activity in 1999.* (Office for National Statistics: 2001)
Report 29	Office for National Statistics Survey Control Unit. *Report on GSS Survey Activity in 1999.* (Office for National Statistics: 2001)
Report 30	Eileen Goddard. *The value and feasibility of a national survey of drug use among adults in the United Kingdom*. (Office for National Statistics: 2002)
Report 31	Office for National Statistics Quality Centre. *Report on GSS Survey Activity in 2000*. (Office for National Statistics: 2002)
Report 32	Peter Sharp. *Gross Domestic Product: Output approach (Gross Value Added) – Revised*. (Office for National Statistics: 2003).

All of these reports are available in pdf format from the National Statistics website

http://www.statistics.gov.uk/methods_quality/publications.asp