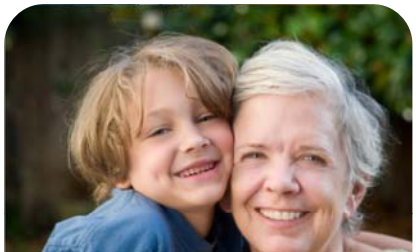
Office for National Statistics

Population Trends

Autumn 2010

Edition No: **141**
Editor: **Chris W Smith**
Office for National Statistics



141

ISSN 2040–1590

A National Statistics publication

National Statistics are produced to high professional standards set out in the Code of Practice for Official Statistics. They are produced free from political influence.

About us

The Office for National Statistics

The Office for National Statistics (ONS) is the executive office of the UK Statistics Authority, a non-ministerial department which reports directly to Parliament. ONS is the UK government's single largest statistical producer. It compiles information about the UK's society and economy, and provides the evidence-base for policy and decision-making, the allocation of resources, and public accountability. The Director-General of ONS reports directly to the National Statistician who is the Authority's Chief Executive and the Head of the Government Statistical Service.

The Government Statistical Service

The Government Statistical Service (GSS) is a network of professional statisticians and their staff operating both within the Office for National Statistics and across more than 30 other government departments and agencies.

Contacts

This publication

For information about the content of this publication, contact Chris W Smith, editor

Tel: 01329 444683

Email: population.trends@ons.gsi.gov.uk

Other customer enquiries

ONS Customer Contact Centre
Tel: 0845 601 3034
International: +44 (0)845 601 3034
Minicom: 01633 815044
Email: info@statistics.gsi.gov.uk
Fax: 01633 652747
Post: Room 1101, Government Buildings,
Cardiff Road, Newport, South Wales NP10 8XG
www.ons.gov.uk

Media enquiries

Tel: 0845 604 1858
Email: press.office@ons.gsi.gov.uk

Copyright and reproduction

Contents

Available from 23 September 2010 at: **www.statistics.gov.uk/populationtrends/ptissue/**

In brief

Progress on the 2011 Census

Planning for the census is proceeding apace. Most of the key changes to the design and systems arising from the lessons learned from the 2009 Rehearsal (see Neil Townsend's article in *Population Trends 140*) are in place. The top supervisory level of the local field force –the 150 or more area managers– are all in post, and each is being supported by new teams of recently appointed local community advisors, and managed by a team of regional managers, based at Census HQ at Titchfield. Recruitment for the foot soldiers –the teams of field staff who will deliver census questionnaires to addresses which will not receive questionnaires via Royal Mail (mainly communal establishments) and who will collect those census questionnaires not returned by post or completed on line– will start in January.

The residential address check, which employed 350 checking staff over a period of four months was completed on target in August this year. The address check has covered 3.6 million residential addresses in England and Wales (around 15 per cent of the total list) targeted on those areas where we had most concerns about the quality of the list. At the same time specially trained staff have visited all 'complex' communal establishments - universities, prisons, army camps etc and thousands of other communal establishments where it was felt that a check was required. As well as finding new addresses and resolving conflicts between different address sources the check has provided the strongest evidence yet that the register will hit (or exceed) its quality targets and provide a high quality framework for running the census.

The classifications for coding data from the 2011 Census have been finalised, and are the largest ever produced for a UK Census. They have been developed by business areas across ONS, with some being developed specifically for the 2011 Census and others being standard ONS classifications or census-specific variations of these. The classifications will be used to code responses to a number of questions, following the data capture operation and prior to ONS receiving the data from its contracted data processing service provider. The topics for which a classification is required for the 2011 Census in England and Wales are:

- country of birth and country of non-UK address (in second residence, address one year ago and visitor questions) (same classification for both)
- citizenship (passports held)
- national identity
- ethnic group
- religion
- language
- occupation
- industry
- National Statistics socio-economic classification (NS-SEC)

Consultation on census outputs

ONS launched its programme of consultation on the requirements for statistical outputs with a series of national roadshow events in October last year, and published its 2011 Census Output Strategy in December 2009. See: www.ons.gov.uk/census/2011-census/consultations/open-consultations/ 2011-output-consultation---main-statistical-outputs/2011-census-outputs-strategy.pdf

This was followed by a formal online consultation that ran from14 December 2009 to 16 April 2010 to allow census users to contribute to the design of the main body of outputs. Users were asked to comment on the specification whilst being free to add any additional requirements they wished including, new tables, geographies, derivations and classifications. Feedback from the consultation has been positive and 134 responses were received.

The Census Outputs team is analysing the responses and specifying the tabulations for submission to a second round of consultation later this year to confirm that users are content. Issues being considered include:

- disclosure control
- new delivery mechanisms - including the degree of flexibility that users will be offered
- the level of detail requested by users
- a separate set of evaluations at each geographical level for which outputs will be produced.

New data available from the ONS Longitudinal Study

The annual refresh of vital event registrations linked to the Office for National Statistics Longitudinal Study (ONS LS) is now complete. The ONS LS was set up in the 1970s to meet the need for better data on mortality and fertility. Since then it has been used to address a wide range of research questions including studies of social mobility, ageing and migration. The refresh means that researchers using the ONS LS have access to an extra year of vital event information for their research. Information on vital events for 2008 and on cancer registrations for 2007 are available.

The ONS LS contains linked census and vital event data for one per cent of the population of England and Wales. Information from the 1971, 1981, 1991 and 2001 Censuses has been linked together, along with information on events such as births, deaths and cancer registrations. At each census, data on more than 500,000 sample members have been included. During the 37 years of the study, more than one million people have been recorded in the sample at some point.

The ONS actively promotes wide use of this data source while maintaining the confidentiality of the individuals in the sample. The records available for analysis are anonymised. Any data that is not aggregated or disguised is held at ONS sites and is only accessible from secure areas known as the Virtual Microdata Laboratory. Because of this, Support Officers are available to help with data extraction and use. Work to link data collected at the 2011 Census is well underway. This involves ONS working closely with the NHS Information Centre in order for planning and preliminary work to run smoothly. These data will be available to researchers in autumn 2013.
For more information on the ONS LS and who to contact if you have any queries please visit:
www.ons.gov.uk/about/who-we-are/our-services/longitudinal-study
Government and non-academic users: Tel: 44 (0)1633 455 844 Email: maus@ons.gsi.gov.uk
Website: www.ons.gov.uk/about/who-we-are/our-services/longitudinal-study

Academic users: Email: celsius@lshtm.ac.uk Website: www.celsius.lshtm.ac.uk

Pension Trends: updated topics released online

In April and June 2010 the Office for National Statistics published two updated chapters of Pension Trends: Chapter 2: Population change and Chapter 3: Life expectancy and healthy ageing.

Chapter 2 shows that as the UK population ages, the old age dependency ratio –which measures the number of people of State Pension Age (SPA) and over for every 1,000 people of working age– is projected to rise. The ratio was steady at around 300 from the mid-1970s to 2006, but reached 310 in 2008 and, in the absence of any increases in SPA, would be expected to reach 495 by 2051. With the increases in the state pension age which are due to take place between 2010 and 2046 under current legislation, the old age dependency ratio is expected to be 343 in 2051.

Chapter 3 shows that cohort life expectancy at SPA will decline for women over the next decade. Between 2021 and 2051 it is expected to level off for both sexes as planned increases in SPA match projected increases in life expectancy. Cohort life expectancy at SPA between 2021 and 2051 is projected to be around 25 years for women and under 23 years for men.

A related question, explored in Chapter 3, is whether longer life expectancy is associated with longer healthy life expectancy in old age. In 2006, UK men at age 65 had 17.2 years of period life expectancy and 12.9 years of healthy life expectancy, compared with 19.9 years and 14.5 years respectively for UK women. The chapter looks at differences in period life expectancy and healthy life expectancy by key characteristics such as social class, indicators of multiple deprivation, country and region and examines the effects of health risks such as smoking and obesity.

Pension Trends releases are available at: www.statistics.gov.uk/pensiontrends/

For more information on Pension Trends email to: pensionsanalysis@ons.gov.uk

Migration Statistics Annual Report

In 2009 the Migration Statistics Annual Report was introduced as an annual cross-government compendium on migration statistics. Regrettably, it will not be possible to produce an equivalent publication in 2010. Instead ONS will publish a wider range of information in the statistical bulletin accompanying its Long-Term International Migration 2009 release on 25 November 2010.

Quarterly Long-Term International Migration Estimates

In August 2010 ONS introduced provisional quarterly Long-Term International Migration (LTIM) estimates. The latest data cover the calendar year 2009 and are included in the Migration Statistics Quarterly Report, available at: www.statistics.gov.uk/statbase/Product.asp?vlnk=15230

Previously the provisional estimates of long-term international migration were based on data from the IPS only. However, LTIM supplements the IPS estimates with data on asylum seekers, people whose length of stay changes from their original intentions and additional information on international migration to and from Northern Ireland. LTIM therefore offers a complete picture of long-term international migration into and out of the UK.

The Registrar General's review of Scotland's population

Scotland's population continued to rise in the year ending 30 June 2009 to a total of 5,194,000, an increase of around 25,000, or half a per cent, over the previous year.
Speaking about the publication of *'Scotland's Population 2009 – the Registrar General's Annual Review of Demographic Trends'*, Registrar General for Scotland Duncan Macniven said:

"There is little, if any, evidence that the recent recession has affected Scotland's total population. There were just under 1,000 fewer births in 2009 than in 2008, but there were still 1,200 more births in 2009 than in 2007. The steady decline in the number of deaths continued. Almost 22,000 more people moved to Scotland than left the country in the year ending June 2009, slightly more than in the previous year. This was the third-highest movement of people into Scotland since current records began 60 years ago. So, despite the challenging financial climate, people are finding Scotland an attractive place to live and raise children. But there is another side to the story. While the number of deaths has continued to fall and life expectancy has increased in every local authority area, the life expectancy of men and women in Scotland is still lower than the life expectancy of people in the rest of the UK and the European Union (except the East European member states). And there are major inequalities of life expectancy within Scotland. For the average man in North and East Glasgow, life expectancy is 8 years shorter than the average in neighbouring East Dunbartonshire. For women, the difference is around 6 years."

Key points in the publication include:

Population

Scotland's population rose to 5.19 million in 2009, an eighth consecutive annual increase, mainly due to more people moving to Scotland than leaving. Current projections suggest Scotland's population will rise to 5.54 million by 2033 and the number of people aged 60 and over will increase by 50 per cent, from 1.17 million to 1.75 million. In the 10 years from 1999 to 2009 the council areas which had the highest population increases and reductions were:

- West Lothian – up 10 per cent
- East Lothian – up 9 per cent
- Perth and Kinross – up 8 per cent
- Inverclyde – down 6 per cent
- Eilean Siar, East Dunbartonshire and West Dunbartonshire – down 4 per cent.

Migration

In the year to 30 June 2009:

- 45,400 people came to Scotland from the rest of the UK
- 41,300 people left Scotland for other parts of the UK.

This increased the population by around 4,100 people.

In the year to 30 June 2009:

- 42,700 people came to Scotland from overseas
- 25,200 people left Scotland to go overseas.

This increased the population by around 17,500 people, a record high.

Births

There were 59,046 births registered in Scotland in 2009, two per cent fewer than in 2008. The number of births had increased in each of the previous six years. 86 per cent of mothers who gave birth in Scotland in 2009 were born in the UK, including 76 per cent who were born in Scotland. Six per cent of mothers were born in the European Union (EU), including 3.5 per cent from the countries which joined the EU in 2004 (such as Poland).

Deaths

There were 53,856 deaths registered in Scotland in 2009, the lowest number since 1855 (when civil registration started). The main causes of deaths were:

- cancer for 15,187 deaths (28 per cent of all deaths)
- ischaemic (coronary) heart disease, 8,274 deaths (15 per cent of all deaths)
- respiratory system diseases (such as pneumonia), 7,125 of deaths (13 per cent of all deaths)
- cerebrovascular disease (stroke), 4,906 deaths (9 per cent of all deaths).

Deaths from coronary heart disease have fallen from 29 per cent in 1981 to 15 per cent in 2009, but the percentage of deaths due to cancer has risen from 22 per cent to 28 per cent.

Life expectancy

Life expectancy in Scotland has increased from 69.1 years for men and 75.4 years for women born around 1981, to 75.3 years for men and 80.1 years for women born around 2008. Despite recent improvements, Scottish men and women have poor life expectancy compared to most of the EU, about four years lower for men, and almost five years lower for women, when compared to the countries where life expectancy is highest.

Marriages and civil partnerships

There were 27,254 marriages in Scotland in 2009. This includes 6,664 marriages (24 per cent) where neither the bride nor groom lived in Scotland, but does not include people living in Scotland who marry elsewhere. Just over half of all marriages (52 per cent) were civil ceremonies, carried out by a registrar, compared to just under one-third (31 per cent) in 1971. Most religious marriages (6,143) were carried out by Church of Scotland ministers, with clergy from the Roman Catholic Church carrying out 1,788 marriages. Celebrants from the Humanist Society of Scotland, authorised to carry out marriages since 2005, officiated at 1,544 marriages.

In 2009 there were 498 civil partnerships, 219 male couples and 279 female couples. There were 10,371 divorces and 24 dissolutions of civil partnerships.

October 2010: ONS Migration Statistics Improvement Programme development seminars

As announced in *Population Trends 140*, ONS plans to hold seminars in October 2010 to explain developments planned for Phase 2 of the Migration Statistics Improvement Programme (MSIP).

The objectives of the seminars are:

- to provide stakeholders with an overview of the plans and progress of the cross-government MSIP and their implications for population estimates

- to gather feedback from stakeholders on the plans for Phase 2 of the MSIP to feed into the programme
- to meet the requirement of the UK Statistics Authority Code of Practice for Official Statistics to 'engage effectively with users of statistics to promote trust and maximise public value'

The seminars are held on the following dates and locations:

Tuesday 05 October 2010	Leeds, Leeds Armouries
Wednesday 06 October 2010	Coventry, HMRC, Sherbourne House
Thursday 14 October 2010	London, HMRC, Somerset House
Tuesday 19 October 2010	Newport, South Wales, ONS, Government Buildings

Attendance is free of charge. If required, documentation in Welsh will be made available at the Newport seminar. A light buffet lunch and refreshments will be provided. A detailed agenda will be available ahead of the seminars.

To register for an event, please download and complete a registration form via this link: www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-stats-improve-prog/comm-stakeholders/msip-roadshows-2010/detailed-announcement/index.html

Alternatively, registration forms can be requested by email from imps@ons.gov.uk or by calling 01329 - 444 584

Quick online access to older Population Trends articles

Readers interested in locating previously published articles may like to know that a searchable database is available. This online article search facility covers all ONS journals. To find an article it is possible to do a text search for keywords, journal title, article title, author's name, issue number, and publication year. All articles published in Population Trends since Winter 1997 (issue nr. 90) are available online. Using this free search facility, pdf files can be downloaded for each article. To use the service, go to: www.statistics.gov.uk/cci/articlesearch.asp

Population Trends: readers' views invited

At ONS we continually strive to maintain the quality of Population Trends as an important demographic journal. The views of our readership are important to us and we would welcome any comments and suggestions you have about the future scope and direction of the journal to ensure it remains fresh and pertinent while maintaining the high standards expected by our readership.

Please email your comments and suggestions to: population.trends@ons.gov.uk

Readers are also reminded that we always welcome submission of papers from external authors that are appropriate to the scope of the journal.

The ESRC Centre for Population Change – working in partnership with ONS and GROS

Jane Falkingham

Director ESRC Centre for Population Change, University of Southampton

This October marks the first anniversary of the formal launch of the ESRC Centre for Population Change. This issue of Population Trends includes a number of articles based on research within the Centre, highlighting the breadth of research taking place within the Centre and its relevance for researchers in academia and national and local government.

The ESRC Centre for Population Change (CPC) brings together experts from across the social sciences to carry out research that will lead to a better understanding of the key drivers of population change and the implications for economic welfare and social support at the national, local, household and individual level. The CPC is a joint initiative between the University of Southampton and a consortium of Scottish Universities, led by the University of St Andrews, in partnership with the Office for National Statistics and the General Register Office for Scotland[1].

The Centre is carrying out research in the thematic areas of:

1. Dynamics of fertility and family formation;
2. Household change and living arrangements across the life course;
3. The demographic and socio-economic implications of national and transnational migration;
4. Modelling population growth and enhancing the evidence base for policy.

Within each of the four themes a series of research projects is addressing some of the critical scientific research questions. The choice of these themes reflects the critical role played in recent population change by:

- Low fertility, with couples now having, on average, fewer than 2.1 children, below the level needed for the population to replace itself. A key question is whether this trend will continue or whether fertility will rise and what, if anything, governments can or should do to influence this.

- The emergence of 'new' forms of intergenerational relationships and living arrangements, such as 'living apart together', ‘boomerang children’, that is, adult children returning to the parental home in their 30s and 40s, and increasing childlessness. Improvements in longevity and rises in the average age at which women give birth and the age at which children subsequently leave home have also resulted in more people, particularly women, being faced with the double burden of caring for both their own (adult) children and their parents, the so-called ‘sandwich generation’. Again these new behaviours raise important issues for policy makers. More information is required on how widespread these behaviours are and the factors associated with them.
- Increasing levels of migration both internally and between countries. Of particular note is the increase in the number of non-UK EU migrants settling in the UK, many of whom are of childbearing age. One key question is whether these migrants will choose to stay and raise a family here or whether they are just temporary migrants who will return home. This has important consequences both for the UK in terms of its demographic structure, labour market and social transfers and for the country of origin, where the economy as a whole, as well as individual families may benefit from remittances but where the disruption of traditional family networks may leave older people vulnerable.
- Individual and population ageing. The next 20 years will see an increase in the proportion and absolute number of older people in the population. This presents challenges and opportunities for society; challenges in terms of providing economic support and health and social care and opportunities in terms of more people living a healthy and productive life and for longer.

A hallmark of the Centre’s research is a focus on the dynamic interconnections between the different demographic processes and the implications for policy and the community in the context of an ageing society. Ranging from in-depth qualitative studies of the processes underlying individual demographic behaviour through to complex statistical and economic modelling, the Centre’s research brings methodological innovation to the study of population change in the UK and will enhance the evidence base both for improved population projections and for policy formulation. Partnerships with ONS and GROS ensure that the work undertaken is relevant to national and local government.

The papers by CPC researchers included in this issue of Population Trends were chosen to showcase both the range of research topics and the variety of datasets being used within the Centre. The first article by Máire Ní Bhrolcháin, Eva Beaujouan and Ann Berrington presents the initial results from a new combined dataset bringing together data from the General Household Survey for the period 1991-2007. The paper examines time trends in family intentions in order to improve our understanding of recent trends in fertility. The research challenges the idea that the trend toward later childbearing is the result of active ‘postponement’ but argues that the whole environment of opportunities and incentives for childbearing has been changing, resulting in a progressive increase in the age of mothers at all parities.

This is followed by an article by Irene Mosca and Robert Wright who use HESA data to examine the nature of national and international graduate migration flows in the UK. The analysis suggests that migration is a selective process with graduates with certain characteristics having considerably higher probabilities of migrating both to other regions of the UK and abroad.

Moving further up the life course, Athina Vlachantoni investigates the demographic characteristics and economic activity patterns of carers over the age of 50 in England using data from the English Longitudinal survey of Ageing (ELSA). The results suggest that the nature of care provision changes at different points of the lifecourse, and that it is quite a different experience for men and women. In particular, the research sheds new light on the so-called ‘round-the-clock’ carers, who make up just over one-fifth of all carers aged 50 and over.

Maria Evandrou, Jane Falkingham and Marcus Green then explore the factors associated with migration among people aged 50 and over using data from the British Household Panel Study over the period 1991–2007. Previous research has distinguished three distinct types of migration at older ages: ‘amenity’ related or retirement migration undertaken to improve quality of life; disability or health-related migration, undertaken to move closer to kin; and finally moves into institutions towards the end of life. The paper highlights that as divorce and remarriage become more common in later life, a fourth category of ‘relationship driven migration’ may need to be added to this typology.

Finally, Guy Abel, Jakub Bijak and James Raymer from the CPC modelling strand compare official population projections with Bayesian time series forecasts for England and Wales that allow of the inclusion of uncertainty in a consistent manner. The results lend support to the idea that the methods currently used to produce population projections should be modified to include a greater emphasis on specifying uncertainty, allowing more informed decisions to be made by population planners and policy makers.

1 CPC is directed by Professor Jane Falkingham and co-directed by Professor Maria Evandrou (University of Southampton) and Dr Elspeth Graham (University of St Andrews). Its founding co-directors included Professor Paul Boyle, who from 1 September 2010 has been appointed Chief Executive and Deputy Chair of the Economic and Social Research Council, and Professor Sue Heath who has recently taken up a new post at the Morgan Centre for the Study of Relationships and Personal Life at Manchester University.

Stability and change in fertility intentions in Britain, 1991-2007

Máire Ní Bhrolcháin, Eva Beaujouan, Ann Berrington

University of Southampton

Abstract

The very low fertility experienced in several European countries in recent decades in the presence of higher intended family sizes has renewed interest in fertility intentions data. While the overall level of childbearing in Britain over the past few decades has remained relatively stable and high in comparison with many other European countries, we have seen sizeable increases in the age at which childbearing starts. This study uses data from the 1991 to 2007 General Household Surveys to examine trends in family intentions data in an attempt to arrive at a better understanding of these recent fertility developments. First, time trends in intended family size are compared with trends in observed fertility. Next, aggregate changes in intentions regarding the level and timing of fertility across the life course for cohorts are investigated together with the extent to which these aggregate intentions are matched by the subsequent childbearing of cohorts. Finally, both change across the life course and uncertainty in family intentions are examined. We conclude by discussing what these findings might tell us about contemporary reproductive decision making.

Contents

Introduction

Fertility in Britain in the last decades has displayed both stability and change. Stability is seen in the narrow range of 1.62 to 1.95 within which the total fertility rate (TFR) -an annual summary indicator of hypothetical family size if current birth rates were to persist- has been fluctuating since the early 1970s. Substantial change has been occurring, on the other hand, in the timetable of family formation. For three decades, women have been having children at progressively older ages than customary in the recent past with the result that the age-standardised mean age at childbearing is, at 29.3 years, now higher than at any time since 1938, when records of age at birth first became available in vital statistics. Britain's combination of several decades of relatively stable, low fertility with a strong trend to later childbearing is shared with other countries of northern and western Europe[1]. But, along with France and some Scandinavian countries, it has maintained a level of fertility well above that of some central and southern European countries. For example, around 2006, the TFR in countries of western and northern Europe (excluding the German-speaking countries) was between 1.7 and 2.0, while most other regions of Europe displayed TFRs below 1.5. Almost all European countries have seen a rapid increase in the age at first birth, but inter-country differences remain. For example, in 2006, mean age at first birth in Spain was around 29 years, while in Russia childbearing was much earlier, with an average age at first birth of 24 years[2].

While the overall level of fertility in Britain has, in recent decades, been relatively stable, there have nevertheless been some short to medium-term trends. During the 1990s the TFR declined steadily from 1.84 in 1991, reaching a low of 1.62 in 2001. The TFR then turned up over several years, reaching 1.95 in 2008, an increase of 0.33 births per woman (20 per cent) over 2001. Most (69 per cent) of the rise of 0.33 children per woman from 2001-2008 was due to rising birth rates of women aged 30 and over, continuing a trend in place since the late 1970s. A further 25 per cent is due to women aged 25-29, reflecting a slight turnaround in the birth propensity at this age. Women in their early twenties account for a small part of the increase, with little indication as yet of a move away from the relatively low rates reached at these younger ages. Rising fertility has also been observed in the new millennium in several other European countries, the causes of which are yet to be established[3]. Suggested explanations include an end to the move to later childbearing, economic growth, pronatalist and other family policies, and higher immigrant fertility. Recent research on the role of migrant childbearing in the UK shows that, while the overall proportion of births to foreign born women has increased significantly, the rise in the overall UK TFR is mainly due to increases in the fertility of UK-born women[4].

The present study examines time trends in family intentions data in an attempt to arrive at a better understanding of recent trends. Fertility intentions data have long been studied in demography and can be deployed in several ways. Early investigators had recourse to such data as an input to population projection assumptions. But optimism in that respect gave way to considerable scepticism, with the accumulation of evidence that aggregate fertility intentions tend to reflect current conditions rather than future prospects[5 6 7]. The very low fertility experienced in several European countries in recent decades has renewed interest in intentions data, as mean intended or desired family size has been found to be above period estimates of family size[8]. This has been interpreted in several ways. The difference between the current TFR and desired family size is seen by some as signalling a likely future upturn of period fertility levels to reach expressed preferences, by others as reflecting the opposite—that declared intentions will ultimately decline to

reflect current actual fertility levels—and alternatively as reflecting a disjunction that could continue indefinitely in low fertility societies[9]. The apparent shortfall of actual compared with desired fertility has also been viewed as resulting from barriers to the achievement of desired family size, and thus as the rationale for family-friendly policy intervention[10]. This latter perspective has prompted several cross-national investigations of fertility intentions in a European context[11].

Beyond issues of future prospects and policy, fertility intentions may contribute to enhancing understanding and interpretation of recent trends in fertility[9] and that is the purpose of the present study. The shift in the timetable of childbearing in recent decades has been substantial but remains to be fully documented and explained. The phenomenon is widely referred to as postponement though doubts have been expressed as to whether that term is either accurate or useful as a description of the underlying behaviour giving rise to changing tempo[12]. In this article, we look to fertility intentions data from the General Household Survey, in an attempt to clarify recent change in family building patterns, updating a previous analysis of the same data to 2002[13]. Our study has four components. We start by looking at time trends in fertility intentions and then assess the accuracy of childless women's fertility intentions in the aggregate. We go on to consider the variation in fertility intentions across the life-course, and finally examine the extent and patterns of uncertainty expressed in answer to questions on fertility intentions. We conclude with some reflections on what these data suggest about the process of reproductive decision making.

Data and definitions

Data

The data used are from a combined file GHS dataset for the period 1991-2007. Only female respondents are asked to report a fertility history and also to provide information about their intended fertility; and so the present analysis is confined to women.

During the earlier part of the period covered by the present study, GHS fieldwork was on a financial year basis. Hence what we refer to as the 1991 GHS round took place between April 1991 and March 1992. In 2005, the survey reverted to a calendar year basis[14]. In a further change, a longitudinal structure was initiated in 2005. As a result, three quarters of households and persons interviewed in 2006 and 2007 were re-interviews. Repeat interviews in 2006 and 2007 are omitted from the present analysis, and as sample numbers are therefore small in 2006 and 2007, the years 2005-07 are combined in all calendar year analyses. The tables and analyses presented here are weighted for the years 1996 to 2007 when survey weights were available for the GHS. Weights have been scaled by the average weight for the year in question.

Official statistics on aggregate fertility trends are generally published separately for England and Wales and for Scotland. However, the GHS covers the population of Great Britain (England, Wales and Scotland). When referring to national statistics not available for GB, we cite England and Wales figures and indicate this in the text. In general, any discrepancy with the true GB figure will not be substantial.

Measures of intentions

Several measures of intentions are used in the present article: whether a (further) birth is intended, intended family size, and, among those intending a (further) birth, the expected age at next birth.

The question wording and sequence is given in **Box one**. The question on intention regarding a (further) birth combined with a probe for “don’t knows”, has the categories: “yes”, “probably yes”. “no”, “probably no”, and “don’t know”.

Intended family size is measured as follows: (a) the stated intended family size of women answering “yes” or “probably yes” to the question on whether they expected to have any (more) children; in the small proportion of cases where the woman had already had more than the recorded intended number, her actual parity was used; (b) where women answered “no”, “probably no” or “don’t know” to the intentions questions, the number of births that respondents had already had, augmented by one if the woman was pregnant at interview. The latter addition was for consistency with those answering “yes” or “probably yes”, who were asked to include the baby they were expecting in their stated intended family size.

Box one Fertility intentions questions asked of women aged 16-49 years in the GHS from 1991 onwards

Ask all women aged 16 to 49

Do you think that you will have any (more) children (after the one you are expecting)? Could you choose your answers from this card.

Yes..... 1

Probably yes..... 2

Probably not 3

No 4

Don’t know..... 5

Ask if respondent answered don’t know above

On the whole do you think...

You will probably have any/more children 1

Or you will probably not have any/more children? 2

Ask if respondent is likely to have more children

How many children do you think you will have born to you in total, including those you have had already who are still alive/ (and) the one you are expecting?

How old do you think you will be when you have your first/next baby (after the one you are expecting)?

Women answering “yes” or “probably yes” to the question on whether they expected to have a (further) birth were asked to state the age at which they expected to have that birth. The expected time to (next) birth is measured as the difference between this age and the respondent’s age at interview.

As Smallwood and Jefferies[13] noted, the GHS questions on fertility intentions are worded so as to elicit realistic answers on their fertility expectations, rather than ideals or desires. With previous authors, we regard intentions and expectations in this area as synonymous.

Time period

To ensure comparability across time, the analysis is confined to intentions data for the period 1991 onwards. Between 1979 and 1990, a sizeable minority of women (nine per cent to 13 per cent) answered “don’t know” to the fertility intentions question. The question format changed in 1991 with the addition of “probably yes” and “probably no” categories, thus reducing the number of “don’t knows”. Also, a question was added to probe “don’t know” answers further (see Smallwood and Jefferies 2003, Annex A)[13]. The additional question results in a re-classification of 71 per cent of the original three per cent of “don’t know” answers to “probably yes” (40 per cent) or “probably no” (31 per cent), thus reducing non-statement to under one per cent of the total.

Parity

Parity, the number of births a woman has had, should ideally be based on the declared number of live births reported in the Family Information section of the GHS. However, a recent study has identified deficiencies in the GHS fertility histories. Murphy (2009) shows that when followed from one GHS round to later rounds, several older birth cohorts record an (intra-cohort) increase in childlessness. For example, 12.8 per cent of women born in 1950-54 were childless at age 40-44, but this figure had risen to 20.7 per cent of the same cohort at ages 55-59[15]. Further analysis has revealed, however, that sizeable proportions of women who were declared childless according to the Family Information section of the survey had reported children of their own living in the household, particularly from the GHS round 2000-01 onward[16]. A revised birth history was therefore constructed by combining the live births declared in the Family Information section of the questionnaire with the birth dates of own children in the household, for the period 1994 on. These revised fertility histories are used in the present article both to define parity at survey and as an estimate of the intended family size of women not expecting any further birth. They are, however, provisional in that a full assessment of the original errors in the fertility histories, and of the accuracy of the reconstruction, remains to be finalised.

Findings

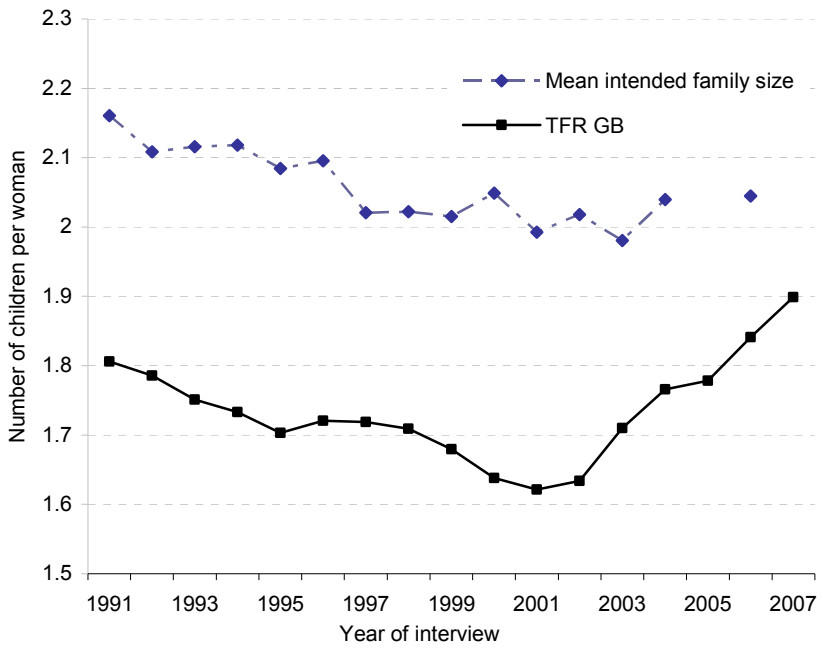
Time trends in intentions

We consider in this section whether and how time-trends in stated intentions are associated with fertility trends. Do intentions data help in anticipating future childbearing trends, or in interpreting recent developments?

On the question of anticipation, the answer is negative. Average intentions in the GHS gave no early warning of the upturn in fertility from 2001 onwards. An intended period mean family size analogous to the period TFR - the unweighted average mean intended family size across single

years of age 16-39 in each year - is shown in **Figure 1** for the period 1991-2005/07, together with the TFR (England and Wales) for the same period (for comparability with vital registration statistics, the intentions data are plotted against year of interview rather than survey year). Mean intended family size ranges between 2.0 to 2.16 children per woman over this 17 year period. Throughout the period it is well above the observed TFR by between 0.3 and 0.4 children per woman to 2001, with the gap narrowing somewhat thereafter. Mean intended family size declines along with the TFR from 1991 on, but what evidence there is of an upturn in aggregate intentions is simultaneous with, rather than predating, the change point in total fertility. Furthermore, the change from 2001-2005/07 in mean intended family size of 0.07 children per woman was just a quarter of the change in total fertility between those dates. These findings are in accord with long-standing evidence that, in the aggregate, fertility intentions tend either to coincide with or to lag, rather than to lead, period trends in fertility[5 6].

Figure 1 Mean intended family size and total fertility rate, Great Britain 1991-2007



Sample: mean intentions based on women aged 16-39.

For comparability with vital registration statistics, the intentions data are plotted against year of interview rather than survey year; for consistency, the combined years 2005-07 are plotted against 2006.

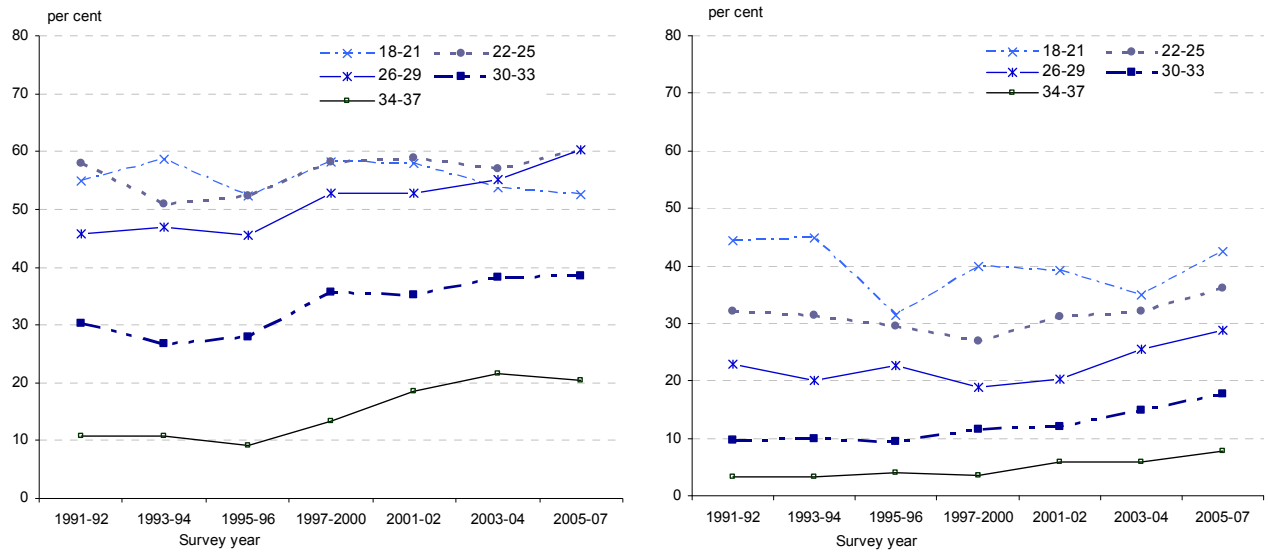
Source: CPC GHS time series datafile

The correspondence between trends in intentions and in aggregate fertility is, however, confined to very broad aggregate measures. The picture alters when we examine a more specific measure of intentions, that is, whether a woman expects to have a (further) birth. This type of question is seen by several investigators as more concrete and immediate, and having a sounder evidential base in relation to validity and reliability, than ultimate intended family size[17]. Furthermore it fits well with established demographic evidence that fertility decisions in low fertility societies are specific by parity, family building decisions being taken one at a time. Unlike average intentions, this

alternative indicator did not, in general, follow the TFR's decline and recovery during the period 1991-2007. Also, trends in the intention to have at least one more child differ somewhat between childless women and those who are already mothers. This is seen in **Figure 2a** and **2b**, which show trends in the proportion saying "yes" to the question on whether the respondent expects to have a (further) birth. From the mid 1990s, the proportion intending at least one further birth shifts upward among childless women in their mid-20s and above[18]. Among women with children, the intention to have a further birth initially declines among younger women, and is constant in older groups, before rising in all groups later in the period. Note that Figures 2a, b give the proportions stating a definite intention to have a birth but that, as we will see later, a substantial minority also express uncertainty about future childbearing. Hence, the complement of the proportions plotted in these diagrams is not the proportion stating that they do not wish to have a (further) birth. In fact, the proportion stating that they do not intend a (further) birth among both childless women and those with children drifts down over the period in all but the youngest age groups, as is seen in **Figure 3a** and **3b**.

Figure 2 Percentage intending to have a (further) child by age at survey and year, Great Britain, 1991/92-2005/07

a) childless women **b) women of parity 1+**



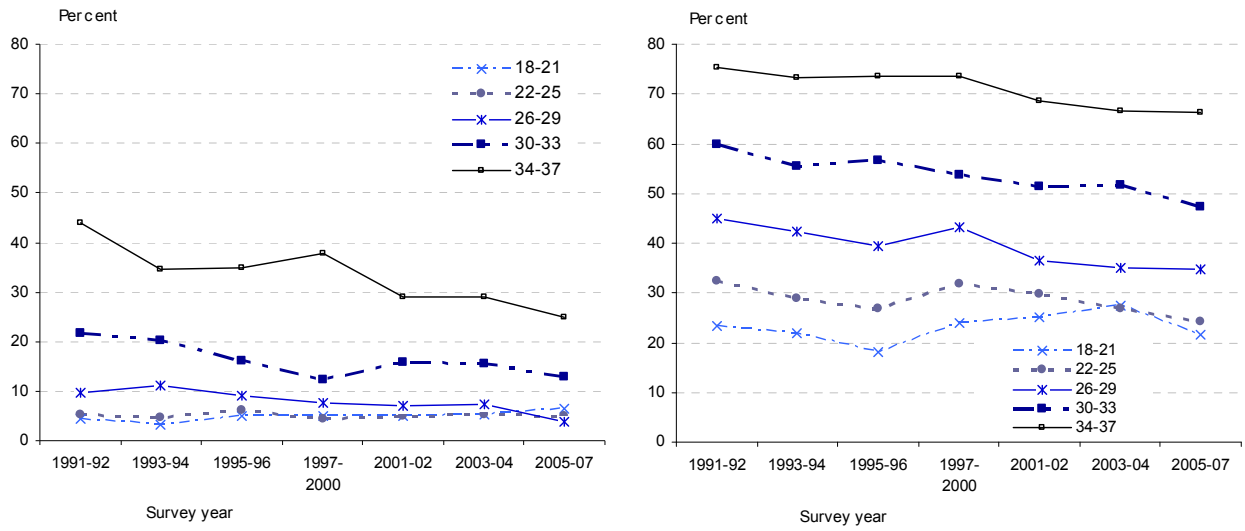
Note: proportions answering "yes" to the first question in Box one.

Source: CPC GHS time series datafile

These findings have implications for the interpretation both of fertility trends and of information on intentions. A process of delaying first birth has, in general, been inferred from time trends in age specific rates and from the rise in the mean age at first birth. That inference goes beyond the evidence, however, in that the move to later childbearing could result from several underlying processes[12]. The intentions data provide direct contemporaneous evidence that confirms recent fertility trends as a process of progressive delay, rather than for example, that the declining rates at younger ages reflected initial decisions against childbearing that were subsequently reversed at later ages.

Figure 3 Percentage intending not to have a (further) child by age at survey and year, Great Britain, 1991/92-2005/07

a) childless women **b) women of parity 1+**



Note: proportions answering "no" to the first question in Box one.

Source: CPC GHS time series datafile

Falling rates at younger ages resulted in an increasing prevalence of childlessness among women in their late twenties and above, as is shown in **Table 1**. For example, 41 per cent of women aged 25-29 in 1991-94 were childless, rising to 56.5 per cent by 2005/07; at ages 30-33 the figure rose from 24.8 per cent to 36.5 per cent ; at ages 26 and above, the rises in proportion childless over the period are statistically significant[19]. However, the growth in childlessness was accompanied by a growing intention among childless women to have at least one child, and there is no evidence of a move away from childbearing per se. As we saw in Figure 3, through the 1990s the proportion of women in their mid-20s and above stating that they did not want a (further) birth was either static or declining. This was true both for childless women of all ages, and for women aged 20 and over who already had children. In all, contemporaneous intentions information supports the widespread interpretation of recent trends as reflecting a continuous process of delaying rather than foregoing motherhood. Thus, birth intentions data appear to be helpful in interpreting current trends, a point to which we return.

Table 1 Childless women by age, per cent, Great Britain 1991- 2007

Age	1991-94 19	95-2000 2	001-04	2005-07
18-21	84.4	84.1	84.0	84.9
22-25	64.0	68.3	69.5	67.4
26-29	41.0	47.0	54.0	56.5
30-33	24.8	30.5	37.7	36.5
34-37	17.7	19.2	22.1	26.3
38-41 14.2		14.8	17.0	18.3

Note: cell sizes (unweighted) are in the range 650 to 2300

Source: CPC GHS time series datafile

The rising intention for at least one birth among childless women indicates that birth intentions data should be interpreted in their temporal context, and viewed against the backdrop of past aggregate trends. Where, as in recent decades, the schedule of childbearing has been shifting up the age range, women of parity zero at later time points in the process are a larger fraction of their age group and thus likely to be less select, in respect of both family size intentions and fecundity, than childless women of the same age at earlier phases of that temporal change[20]. Where, in the presence of such delay, there is little change in the ultimate intention to start childbearing, the outcome is a positive association between proportions childless and intention to have at least one further birth. This is seen in **Figure 4**, which plots, for four-year age groups 22-25 to 34-37, the proportion of childless women in each survey year who expected to have at least one birth (y axis) against the proportion childless at that age in the corresponding year (x axis), during the period 1991-2005/07. At all ages these are directly associated, with correlation coefficients in the range 0.6 to 0.8. The link is particularly strong from the late 20s to the early 30s, the ages at which the greatest change in childlessness has been occurring. During this period, then, the proportion of parity zero women intending to have a birth is closely and directly linked with the proportion of the age cohort who are childless. The association is absent among younger childless women, and is less evident among parous women.

The net result of these offsetting movements is that the proportion of women at each age who either had, or expected to have, at least one child changed little over the period, as is seen in **Table 2**. In 1991-94 between 86 per cent and 89 per cent across age groups either had or expected to have at least one birth, and the range in 2005/07 was 84 per cent -91 per cent. There has, then, been no decline in the combined proportions actually having or intending to have at least one birth. Nevertheless, later childbearing means that a declining proportion have become mothers by a given age: for example, in 1991-94, 87 per cent of women aged 30-33 who had or intended to have at least one child had already had their first birth compared with 72 per cent in 2005/07; the figures are 95 per cent in 1991-94 and 86 per cent in 2005-07 for women aged 34-37. We examine in the next section how far expected and actual levels of progression to first birth coincide at the aggregate level.

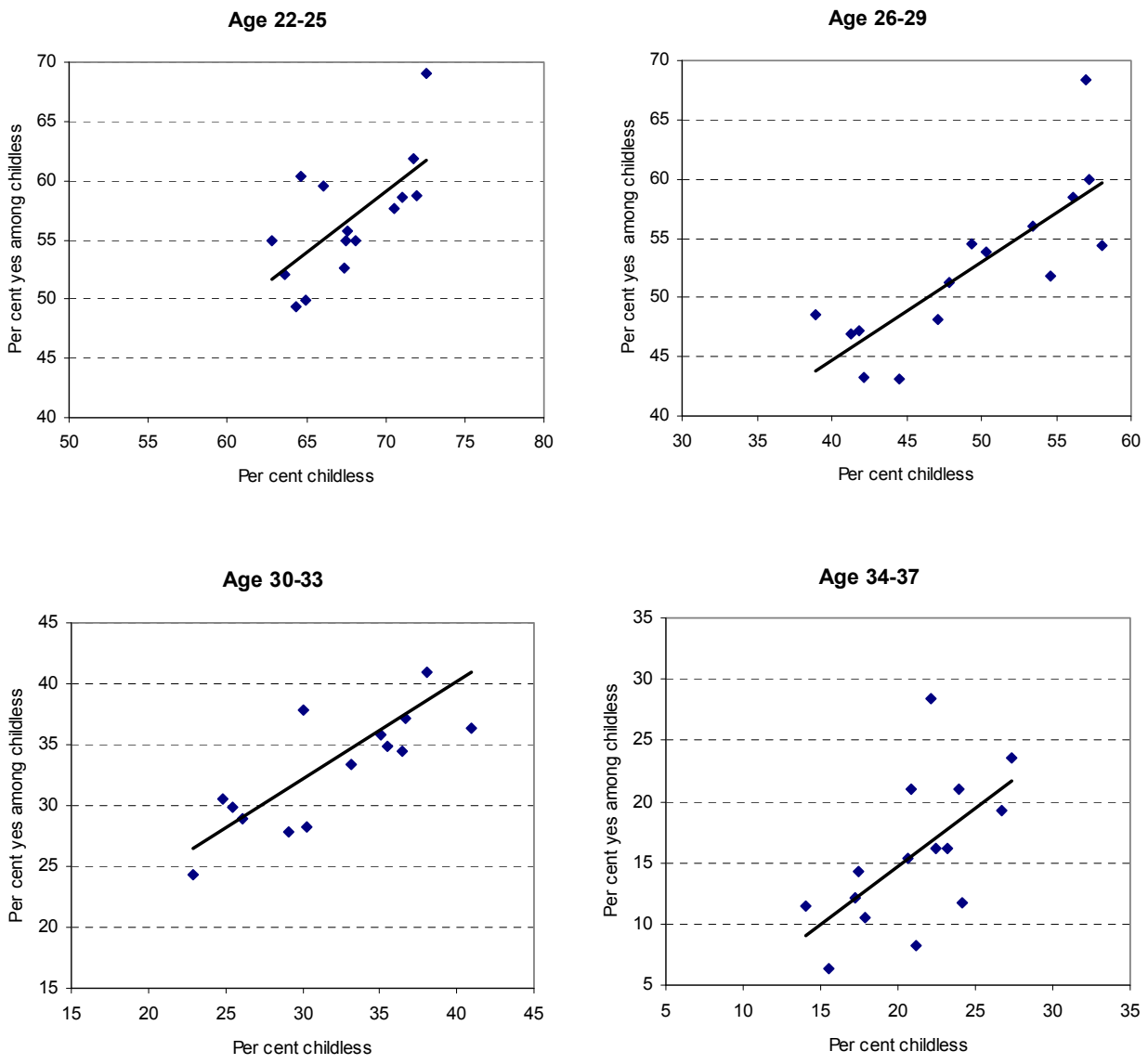
Intentions regarding first birth and its timing

As noted earlier, the dominant influence on aggregate fertility in recent years is the progressively later age at childbearing. This is due largely to later ages at first birth rather than to a slower tempo of progression to higher order births. Between 1974 and 2007 the (age-standardised) mean age at childbearing rose in England and Wales by just over three and a half years, from 26.1 to 29.7. This is attributable entirely to later ages at first birth, the (age-standardised) mean age at first birth having risen by 3.5 years during the period, from 24.0 in 1974 to 27.5 in 2007. Half of this delay took place between 1974 and 1991 (increase in mean age at first birth of 1.7 years), and half subsequent to 1991 (further increase of 1.8 years)[21]. We look here at how far intentional delay, or awareness of a likely delay, is evident in women's stated intentions.

Women who said that they would, or probably would, have a (further) birth were asked in the GHS interview to say at what age they expected to have their next birth. The answers to this question indicate that through the 1990s younger women without children were anticipating a growing delay in childbearing. Childless women aged 18-21 in 1991-92 declared that they expected to have their

first birth 6.2 years later, and the figure had risen to 6.6 years by 2001-02[22]. Among women 22-25, the anticipated wait rose from 4.5 to 4.9 years over the same period; at older ages, the increase is smaller, and in the mid to late 30s absent. The expected waiting time stopped rising in 2001/02 and eased very slightly after the turn of the century (**Table 3**). Thus, the waiting time estimated by younger childless women through the 1990s reflected the upward trend in the time to first birth.

Figure 4 Percentage of childless women who intend to have a birth by proportion of the age group childless in each year, Great Britain1991-2005/07



Source: CPC GHS time series datafile

Table 2a Percentage who have, or expect to have at least one live birth, Great Britain 1991-2005/07

Age	1991-94	1995-2000	2001-04	2005-07
18-21	86.0	88.1	87.0	89.0
22-25	89.0	88.8	88.1	90.9
26-29	88.0	90.0	88.3	91.1
30-33	86.9	89.7	87.4	88.6
34-37	86.3	87.1	87.0	86.0
38-41	86.4	86.9	85.4	84.0

Note: the percentage tabulated is women who already have had or intend at least one birth.

Table 2b Percentage of those in Table 2a who have already had 1+ births, Great Britain 1991-2005/07

Age	1991-94	1995-2000	2001-04	2005-07
18-21	18.1	18.0	18.4	17.0
22-25	40.4	35.7	34.6	35.9
26-29	67.0	58.9	52.1	47.7
30-33	86.5	77.5	71.3	71.7
34-37	95.4	92.8	89.5	85.7
38-41	99.3	98.0	97.2	97.3

Source: CPC GHS time series datafile

Table 3 Mean duration in years before women expect to have their first birth. Childless women stating an expected age, Great Britain 1991/92-2005/07

Age	1991-92	1993-94	1995-96	1997-2000	2001-02	2003-04	2005-07
18-21	6.2	6.3	6.4	6.4	6.6	6.5	6.4
22-25	4.5	4.5	4.7	4.7	4.9	4.9	4.9
26-29	3.3	3.4	3.7	3.5	3.6	3.4	3.4
30-33	2.7	2.9	2.8	3.0	2.9	2.8	2.7
34-37	2.1	1.9	1.9	2.0	2.1	2.0	2.2
38-41	1.6	1.8	1.6	1.6	1.6	1.3	1.5

Source: CPC GHS time series datafile

However, were their expectations accurate? We assess the accuracy of first birth expectations by comparing the stated intentions of childless women by age in 1991-94 regarding both whether and when they expected to have their first child against the actual fertility outcomes of a comparable sample of women ten years older identified from fertility histories collected ten years later in 2001-2004[23]. This comparison allows aggregate (net) accuracy to be evaluated.

The evidence suggests that women in the early 1990s were not particularly accurate in stating either whether or when they would have their first birth. Childless women on average overstate the likelihood of their having their first birth within a specified time. Three in ten childless women aged 18-21 in 1991-94 expected to have their first child within 5 years of the interview, and two in ten of that age group are estimated to have done so[24] (**Table 4**). Zero parity women in their early twenties and above overstate the likelihood of their having a first birth within 5 or 10 years by a

larger margin. For example, while 55.6 per cent of childless women aged 22-25 in 1991-94 expected to have their first birth within 5 years, just 32.4 per cent had done so; comparable figures for women 26-29 are 61.8 per cent and 38.3 per cent. The gap between expected and actual proportions progressing to their first birth within 10 years is also substantial at all ages. If we regard all those who had a birth within 5 and 10 years of the initial date as having intended to have a first birth, we can make an approximate estimate of the fulfilment of the intentions of the childless by age in 1991-94. This ranges from just under half of childless women aged 34-37 who intended a birth within 5 years (12.9/26.3) to 84 per cent (42.9/50.2) of those aged 30-33 who intended a birth within 10 years actually having the birth within the expected time.

Table 4 Percentage of childless women in 1991-94 expecting to have their first child within 5 and 10 years, compared with actual proportions proceeding within those durations estimated from fertility histories of 2001-04, by age in 1991-94, Great Britain

Age	Proceed within 5 years		Proceed within 10 years	
	Expected	Actual	Expected	Actual
18-21	29.9	20.5	76.5	45.5
22-25	55.6	32.4	80.8	57.7
26-29	61.8	38.3	70.6	56.8
30-33	47.8	28.9	50.2	41.9
34-37	26.2	12.9	26.4	17.5

Source: CPC GHS time series datafile

The median times to first birth anticipated by women who expected to have their first birth within 10 years of interview in 1991-94, and a comparable figure for the actual outcome for equivalent women identified in the fertility histories collected at 2001-04 interviews, are given in **Table 5**. Women aged 18-21 specify the likely wait fairly accurately—an expected median of 5.7 years against an actual 5.5 years. Women in their early twenties and above in general anticipate an earlier first birth than actually occurs. These are, however, net figures and at the individual level, the correspondence between expectation and outcome will be less than this; also errors may compensate in different ways in different age groups, and so younger women may be individually no more accurate than are older women. Finally, the figures are conditional on the birth having occurred within 10 years, and as we saw, there is substantial inaccuracy in the aggregate concerning the proportions proceeding to a first birth within that time.

Table 5 Expected and actual median time to first birth among childless non-pregnant women by age in 1991-94, Great Britain

Age	Expected (years)	Actual (years)
18-21	5.7	5.5
22-25	3.9	4.5
26-29	2.6	3.7
30-33	2.0	3.5
34-37	1.4	2.8

Actual medians derived from estimated fertility histories of 2001-04, by age in 1991-94.

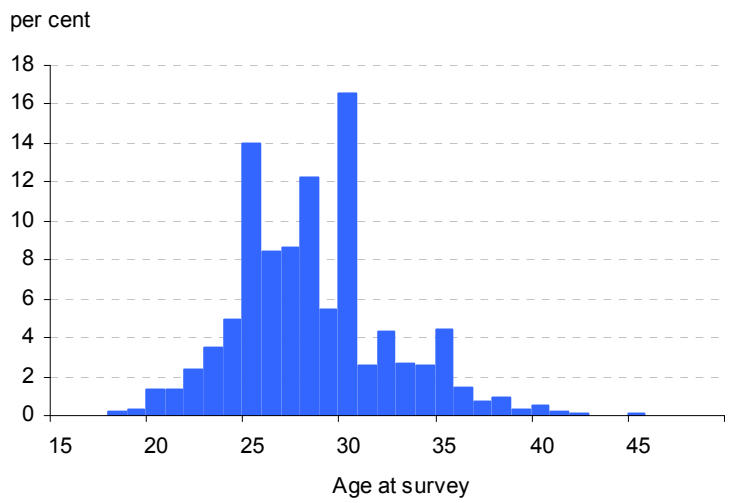
Source: CPC GHS time series datafile

The discrepancies both in the expected versus actual proportions proceeding to a first birth within five or 10 years show that starting a family is far from a precisely timetabled event in the life-cycle. This is evident also from the stated ages at first birth. **Figure 5** shows the distribution of ages at which childless women of all ages in 1991-94 said they expected to have their first birth (period chosen for comparability with Tables 3-5). We see substantial heaping on ages ending in 0 and 5, and to a lesser extent 2 and 8. The distribution has all the classic features of digit preference in the reporting of age in demographic data sources by respondents who either do not know or are unsure of their age; digit preference is found in all periods in these data, as noted by Smallwood and Jefferies (2003)[13]. Heaping in reports of the expected age at first birth is compelling evidence of substantial uncertainty as to both the eventual occurrence and the timing of the first birth.

Life course

We have seen both stability and change in fertility intentions across calendar time. We now examine how far fertility intentions change through the life course The GHS time series allows us to follow up birth cohorts across time periods, thus giving a succession of observations of intentions as cohorts age. As before, we are looking at aggregate change–in this case, the trajectory of cohorts rather than of individuals, and here too both stability and change are in evidence.

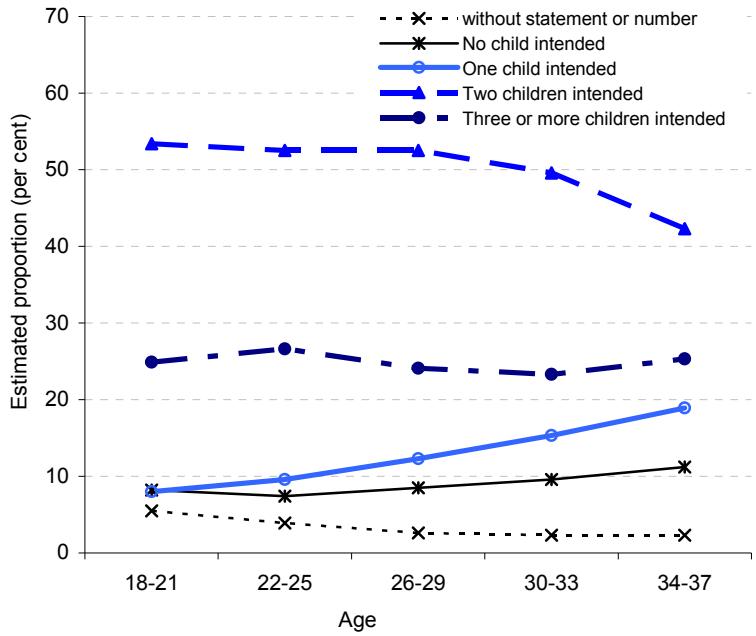
Figure 5 Distribution of expected age at first birth, women of parity zero in1991-94, Great Britain



Among women born in 1968-71 and 1972-75, average intentions declined slightly with age. Intended family size averaged 2.1 children per woman at ages under 25 and registered a decline of 0.1-0.2 by the mid thirties. Much the same is found in other recent studies, though the extent of the reported decline across age is variable with time and place [10 25 26 27 28]. Change in intended mean family size with age is, of course, the result of change in the proportions intending, and ultimately having, specific family sizes. The shift in the distribution of intended family sizes is somewhat more substantial than suggested by the change in average intentions over the life course, in the cohorts considered. We see this from **Figure 6** which summarizes the life course pattern of intra-cohort

change in intentions for specific family sizes for women born in 1968-75. The picture is one of decline in the intention to have two children by age, especially when women reach their 30s, with a corresponding growth in an expected family size of zero or one. An estimated 53 per cent of women intended two children at age 18-21, and this had declined to 42 per cent by age 34-37; the eight per cent intending one child at ages 18-21 had, on the other hand, grown to 19 per cent by ages 34-37. The proportion intending three or more is relatively stable across age[29]. Note that while the focus here is largely on under-achievement of intended family sizes, some women may also have more children than originally intended. In the present aggregate analysis, however, the frequency with which women exceed intended family sizes cannot be determined.

Figure 6 Percentage stating each intended family size by age, cohort 1968-75, Great Britain



Source: CPC GHS time series datafile

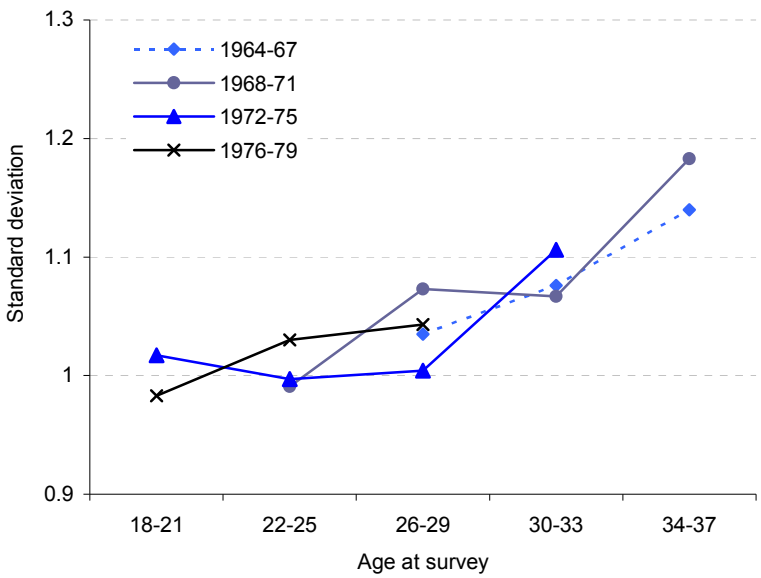
Sample: Birth cohorts 1968-75

Intentions may of course, be unrealistic. In particular, according to stated intentions, 89 per cent of women aged 34-37 in this cohort expected to have at least one child, and thus just 11 per cent expected to be childless. In fact, just 75 per cent of this group had had their first birth at ages 34-37. We can make a rough assessment of the proportion of these women who will eventually have at least one child. If we assume that the proportion who ultimately achieve a birth among those childless at age 34-37 and intending a birth is the same as the estimated 10 year progression from the aggregate 1991-94 to 2001-04 figures in Table 4 above—that is 66 per cent (17.5 per cent / 26.4 per cent)—that would give us an estimate of 16 per cent ultimately childless in this cohort, substantially above the intended 11 per cent, though somewhat below the ONS estimates of 19-20 per cent childless at 45 in the England and Wales cohorts of 1960 and 1963[30].

One result of these shifts is that the variability of intentions grows with age within cohorts, as is seen in **Figure 7**. Women are more uniform in their family size intentions at younger ages but become more differentiated through the life course[31]. The numerical cause is the reduction by the

late thirties in the two-child group and the increase in the one-child category, and to a lesser degree, those stating none. The gradual differentiation of women in their family building intentions as they age mirrors closely the same process in relation to achieved family size, whose variance also increases with age. Towards the end of the childbearing span, then, under recent conditions, women are more diverse in fertility outcomes than their expressions of intentions in their late teens and early twenties. Of course, intentions/expectations are necessarily close to the parity achieved by that age, and intentions at later ages are heavily constrained by the reality of declining fecundity as well as perceived risks, and possibly also normative sanctions stemming from the relative rarity in a low fertility society of childbearing at older ages.

Figure 7 Standard deviation of intended number of births by age, birth cohorts 1964-67 to 1976-79, Great Britain 1991-2007



Source: CPC GHS time series datafile

How certain are women about their fertility intentions?

The question of what may be the origin of apparently shifting intentions across age raises the issue as to how clearly and certainly family building plans are formulated and held at various stages of the life course. It has long been recognised that fertility intentions are subject to considerable uncertainty, and that the inclusion or not of the fertility intentions of the uncertain can significantly affect estimates of mean intended family size[32]. Survey questions on the subject are variable, and the extent of uncertainty associated with fertility intentions varies correspondingly. **Figure 8** shows the distribution by age of answers to the question whether women expected to have a (further) birth, in the GHS 2005-2007. From the base of the graph the answer categories are ordered as follows: no answer, don't know, probably yes, probably no, yes, and no (see also Box one). We combine the first four of these categories and classify them as uncertain. The no answer and don't know groups are a small fraction of those classified uncertain, as can be seen from Figure 8. At ages up to the mid 30s just over three fifths of women give a definite answer on intention for a further birth, and that rises to nearly four fifths by their late 30s. Conversely, right up to the mid 30s, nearly four in ten women give an uncertain answer. The degree of uncertainty declared through the

age range appears quite substantial, given that the question regards not a precise intended family size, but whether a (further) birth is expected.

We see also in Figure 8 that, as previous research has shown, uncertainty declines with age. But the GHS intentions data suggest that after age 30 this is primarily due to women who have already had at least two children. Uncertainty persists well into the thirties for women with fewer than two children. Figure 9 follows the cohort of women born between 1968 and 1975 through time, showing the proportions uncertain at successive ages, and so giving a picture of aggregate change through the life course, specific by parity. Among women with no children, uncertainty rises after age 30 with over half of childless women aged 34-37 being uncertain about their future childbearing. Among those with one child, uncertainty remains relatively high into the 30s and begins to decline only in the late 30s (**Figure 9**). However, uncertainty at the later stages of childbearing is by no means confined to women of below average parity. Among women in their late 30s who declared themselves uncertain about a (further) birth, four in ten had at least two children, a fifth had one child, and just over a third had no children; the childless were, thus, a minority among the uncertain. These data cannot tell us whether the elevated proportions uncertain among childless women in their 30s reflects a selection process, or whether childlessness after 30 of itself gives rise to uncertainty. Another possibility is that it reflects a natural stage in the cycle of family formation[33]. But the level of uncertainty throughout the childbearing years, and its persistence into later ages, suggests that a significant minority of women do not make firm decisions about future childbearing. That a fifth of women at ages 38-41 give uncertain answers suggests a considerable degree of fluidity in orientation to fertility, well into the later years of childbearing.

Figure 8 Distribution of intentions to have a (further) birth by age, Great Britain 2005-07

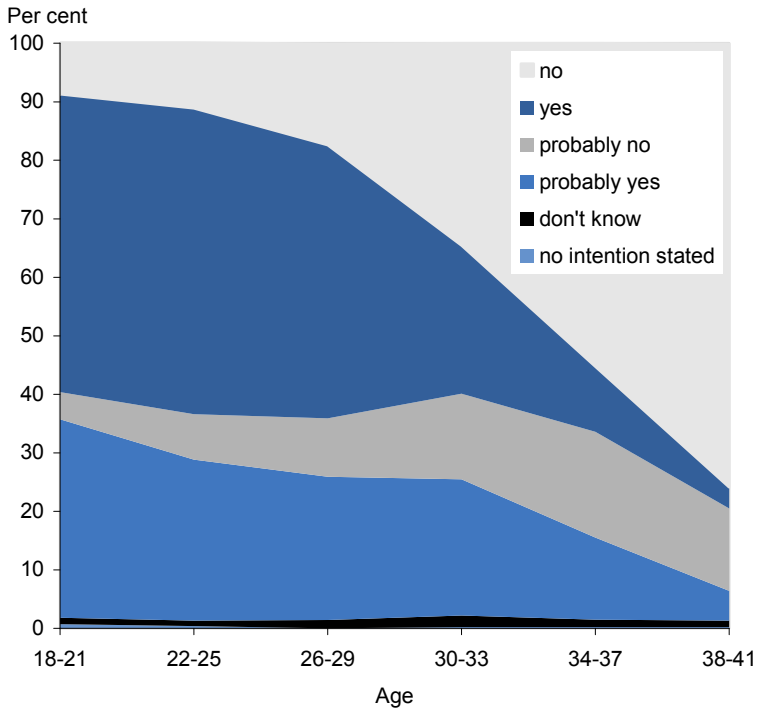
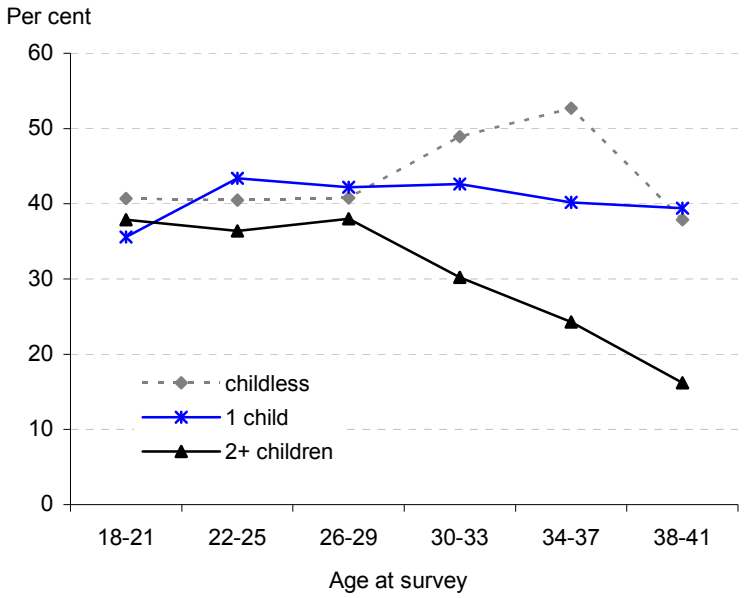


Figure 9 **Percentage uncertain in their fertility intention by parity and age, cohorts born 1968-1974, Great Britain**

Proportion unstated, don't know or probably



Discussion

Mean intended family size during the 1990s tracked overall fertility but did not anticipate the primary change point occurring at the turn of the century. This is much as would be expected from long-standing research showing that trends in aggregate mean intended family size tend either to lag or coincide with overall fertility trends. While average intentions may merely reflect how favourable current conditions are for childbearing, the present study suggests that data on intentions regarding the next birth may be more informative. They might potentially be used as an adjunct to conventional measures of overall fertility, an empirical alternative to tempo adjustment as a guide to the nature and implications of current aggregate fertility trends. However, to establish their validity for this purpose, longer runs of such data would be required, for a variety of temporal contexts, in periods of rising, falling, and static fertility, as well as of differing types of tempo change.

We have seen that women were, through the 1990s, anticipating a growing delay to motherhood, but that their expectations in that respect were inaccurate, in the aggregate, both in respect of the proportions ultimately achieving a first birth and in the timing of the first birth. The term 'postponement' in the context of fertility trends was used originally to refer to a short-term displacement of births in response to short-run conditions adverse to childbearing, such as depression or war, a decline in rates that is compensated by a later, and corresponding, rise[34]. The gap between expected and actual proportions progressing to a first birth, the underestimation of the average delay to first birth among childless women, strong digit preference in the expected age at first birth, and substantial uncertainty about future fertility, all suggest that postponement, in this short-term sense, is not the underlying process driving tempo change. Rather, the whole structure by age of incentives and opportunities for childbearing has been changing, resulting in a

progressive shift up the age range in the fertility schedule[12]. The present data are, however, insufficient to fully illuminate the question as to the type and sequence of decisions that are being made through the life cycle in relation to the reproductive timetable. Large-scale longitudinal data at the individual level would be required to explore and document the behavioural mechanisms at work[12].

There has long been evidence of change across the life course in stated intentions regarding family size. Such change can be seen in different ways. On the one hand, we may view women and couples as having firm fertility intentions when young, and that these change with age due to learning, altered preferences associated with the experience of childbearing, competition with other activities, retrospective rationalisation, and a variety of constraints including fecundity, housing, economic factors, difficulties in partnership formation, partner preferences, as well as period influences[27 28 35 36]. However, the level of uncertainty expressed both in the data analysed here and in other contemporary sources[10 26 37] suggests that many women do not have clear intentions regarding their future childbearing. Both preferences and intentions may be either ill-defined or unformulated for some women over a sizeable part of their reproductive life span, and even if certain of their wishes, their realisation may not be seen as within their control. Change in stated intentions has been found to be related to a variety of individual characteristics and circumstances. But the role of such factors need not necessarily be to modify preferences and intentions, but rather to make them more concrete. Thus, how far the shift in the distribution of intentions from younger to older ages represents true change, how far the crystallization of previously unclear preferences, and how far an adaptation to constraint, is an open question.

Insofar as definite childbearing preferences exist, one constraint on realising these may be establishing a suitable partnership. The presence of a partner is one of the strongest predictors of fulfilment of the intention to have a (further) birth, at the individual level[26 28 38 39 40]. A recent European study found that having a supportive partner was perceived as second only to the mother's health as a circumstance relevant to the decision to have children, being mentioned by over seven in ten respondents. "Lack of right partner" was, in the same study, the leading reason given for having had fewer children than intended[41]. Just under half of childless women in their thirties in the 2005-07 GHS who were either uncertain about their birth expectations or expressed a definite "no", were not in a heterosexual union at the time of interview, compared with slightly over a quarter of those expressing a definite "yes". Partnership issues may thus be contributing both to the childlessness and to the uncertain expectations of this group (though reverse causation may also be at work). Uncertainty about partnership formation is thus a central reason why family formation may be perceived as not under individuals' control, and one potential source of the relatively high levels of uncertainty regarding fertility intentions found in the present study.

The implications at the aggregate level of the substantial uncertainty expressed at all ages are worth considering. Uncertainty pervades the stated intentions of women of all parities. While higher among those with fewer than two children, it features in the answers of a sizeable minority of women with at least two. Between a fifth and a third of women aged 22-37 with two or more children in the period covered by the present study gave uncertain answers on whether they will have a further birth. On this evidence, many women may be keeping their options open, and may be quite ready to react both to change in their personal circumstances and to aggregate period influences, whether favourable or unfavourable to childbearing[33].

Key Findings

- The gap between intended and observed fertility has narrowed since 2000. Mean intended family size during the 1990s tracked overall fertility, but at a level of about 0.3 – 0.4 births per woman higher. Reported intentions did not anticipate the upturn in fertility that occurred around the turn of the century.
- Women were, through the 1990s, anticipating a growing delay to motherhood. But these expectations were inaccurate in the aggregate. Women overstate both the likelihood of their having a first birth, and how soon it will occur.
- Women give very approximate answers when asked when they expect to have their first birth, indicating that starting a family is not a very precisely planned event.
- Among recent cohorts, average intended family size declines somewhat with age. The proportion intending to have two children falls off with age, especially when women reach their 30s, and correspondingly more, as they age, expect a family size of one child or none.
- The level of uncertainty in family building intentions and its persistence into later ages, among women of all family sizes, suggests that a significant minority of women do not make firm decisions about future childbearing.

Acknowledgements

This research is funded by ESRC Grant number RES-625-28-0001. The Centre for Population Change (CPC) is a joint initiative between the University of Southampton and a consortium of Scottish Universities in partnership with ONS and GROS. The findings, interpretations, and conclusions expressed in this article are entirely those of the authors and should not be attributed in any manner to ONS or GROS.

The General Household Survey is conducted by the Office for National Statistics. Access to the data is provided by the UK Data Archive.

The CPC GHS time series datafile was constructed with the assistance of Mark Lyons Amos. We thank the Office for National Statistics and the General Lifestyle Survey Branch for their help in clarifying various data issues.

References

1 See, for example, Demographic Research Special Collection 7: Childbearing Trends and Policies in Europe. Available at: www.demographic-research.org/special/7/

2 Frejka, T and Sobotka, T (2008) Overview Chapter 1: Fertility in Europe: Diverse, delayed and below replacement. *Demographic Research* 19: 15-46.

3 Goldstein, J R, Sobotka, T and Jasilioniene, A (2009) ‘The end of "lowest-low" fertility?’ *Population and Development Review* 35: 663-699.

4 Tromans, N, Jefferies, J and Natamba, E (2009) 'Have women born outside the UK driven the rise in UK births since 2001?' *Population Trends* 136: 28-42.

5 Westoff, C and Ryder, N (1977) The predictive validity of reproductive intentions. *Demography* 14: 431-.453.

6 Lee, RD (1980) 'Aiming at a moving target: Period fertility and changing reproductive goals.' *Population Studies* 34: 205-226.

7 Morgan, SP (2001) *'Should fertility intentions inform fertility forecasts?'* Proceedings of US Census Bureau Conference "The direction of fertility in the United States". Washington, DC: US Census Bureau.

8 Goldstein, J, Lutz, W and Testa, MR (2003) 'The emergence of sub-replacement family size ideals in Europe.' *Population Research and Policy Review* 22: 479-496.

9 Hagewen, K and Morgan, SP (2005) 'Intended and ideal family size in the United States, 1970-2002'. *Population and Development Review* 31: 507-527.

10 Sobotka, T (2009) 'Sub-replacement fertility intentions in Austria'. *European Journal of Population* 25: 387-412.

11 See the special issue of the *European Journal of Population*, volume 25, no 4 on 'Fertility Intentions and Outcomes' edited by D Philipov (2009) and the current European Commission Seventh Framework funded 'Reproductive decision-making in a macro-micro perspective' (REPRO) cross-national study. Available at:www.oeaw.ac.at/vid/repro/

12 Ní Bhrolcháin, M and Toulemon, L (2005) '*Does postponement explain the trend to later childbearing in France?*' Vienna Yearbook of Population Research 83-107.

13 Smallwood, S and Jefferies, J (2003) 'Family building intentions in England and Wales: trends, outcomes and interpretations.' *Population Trends* 112: 15-28.

14 So as not to duplicate cases we omit the first quarter of the GHS 2005 round from the 2005 dataset, as it consists of the final quarter of the 2004-05 survey round. A full year's sample was taken in the second, third and fourth quarters of 2005.

15 Murphy, M (2009) 'Where have all the children gone? Women's reports of more childlessness at older ages than when they were younger in a large-scale continuous household survey in Britain.' *Population Studies* 63 (2) 115-133.

16 Ní Bhrolcháin, M, Beaujouan, E and Murphy, M (forthcoming) 'Childlessness: a further look at cohort estimates based on survey time series data.' Paper to be presented at the European Population Conference, Vienna, 1-4 September 2010.

17 See for example, Casterline, JB and El Zeini, LO (2007) 'The estimation of unwanted fertility'. *Demography* 44: 729-745.

18 Confidence intervals for the estimated proportions intending at least one more birth at the start and end of the period are non-overlapping for childless women in age groups 26-29 and 34-37, when 2-year periods are used, and in all three age groups 26-29, 30-33 and 34-37 when 1991-94 is compared with 2005/07. Among women with children, the confidence intervals are non-overlapping at ages 30 and over, whether 2-year or 4-year periods are used.

19 These figures are based on our current best estimate of the parity of women aged under 40 in the GHS 1991-2005/07, which apply a substantial correction to the anomalies identified by Murphy (2009). However, work is in progress in evaluating these data further, and while our final estimates for these age groups may differ from those given here, any amendments are likely to be minor at ages under 40; at ages over 40+, revisions may be more substantial.

20 The opposite would be true, however, of the later stages of a period of rising fertility, when the childless at any given age would be a more select group than at earlier stages. In general, past childbearing trends are likely to impact on the composition of any parity group with respect to intention, a feature that is additional to any aggregate change in intentions that may be occurring.

21 Figures for more recent years are given in Table 1.7b, Birth Statistics, Series FM1. Data for earlier years supplied by Fertility and Family Analysis Unit, ONS. Figures are ONS estimates of the age at "true" first birth, adjusting vital registration data for the order of births occurring within and outside marriage. The estimates are confined to England and Wales; they are not available for Great Britain. Age-standardisation removes the influence of the population age distribution. See Table 1.7b, Birth Statistics, Series FM1, available at: www.statistics.gov.uk/statbase/Product.asp?vlnk=5768

22 These figures relate to women who stated the age at which they expected to have their first birth. At ages under 26, the rise in expected delay between 1991-92 and 2001-02 is statistically significant.

23 The comparison made is a follow-up of age groups, rather than of individuals. It is carried out as follows, using as an illustration, those aged 18-21 in 1991-1994. The expected occurrence of and wait to first birth of non-pregnant childless women aged 18-21 in GHS rounds 1991-94 is compared with the actual time to first birth of women aged 28-31 interviewed in the 2001-2004 rounds, who, according to the fertility histories collected at 2001-2004, had their first birth within the 9 years and 5 months preceding their 2001-04 interview, or were still childless at 2001-04 interview. The comparison group at 2001-04 is thus defined retrospectively as childless and not pregnant 10 years previously i.e. women who were either of parity zero or a maximum of two months pregnant 10 years before their 2001-04 interview. This assumes that women who were up to two months pregnant would not have reported a pregnancy at interview. The comparison is constructed in this way as women who were pregnant at interview in 1991-94 were asked to state the age at which they thought they would have the next birth after the one they were expecting, and this, for pregnant women, would not have been their first birth.

24 Time to next/first birth is obtained as the difference between the stated expected age at next birth and age at interview. Age at next birth is an integer, and age at survey is, for the purpose, treated as age last birthday rather than exact age. A difference between the two of <=5 and <=10 are classified as within 5 and 10 years, respectively, and so assumes that the age at survey and stated next age are, on average, equal to the integer age plus 0.5 years.

25 Quesnel-Vallee, A and Morgan, SP (2003) 'Missing the target? Correspondence of fertility intentions and behaviour in the US.' *Population Research and Policy Review* 22: 497-525.

26 Berrington, A (2004) 'Perpetual postponers? Women's, men's and couples' fertility intentions and subsequent fertility behaviour.' *Population Trends* 117: 9-19.

27 Hayford, SR (2009) 'The evolution of fertility expectations over the life course.' *Demography* 46: 765-783.

28 Liefbroer, AC (2009) 'Changes in family size intentions across young adulthood: A life course perspective.' *European Journal of Population* 25: 363-386.

29 Similar within-cohort patterns are evident in other sources, though details vary; see for example, Sobotka (2009) Figure 6, note 10.

30 ONS Birth Statistics 2008, Table 10.5. Available at: www.statistics.gov.uk/statbase/Product.asp?vlnk=5768

31 This is consistent with recent findings for the Netherlands, and is reportedly true also of men. See Liefbroer (2009), note 28.

32 See for example, Morgan, PS (1981) 'Intention and uncertainty at later stages of childbearing: the United States 1965 and 1970.' *Demography* 18 (3) 267-285.

33 Morgan, PS (1982) 'Parity-specific intentions and uncertainty: the United States, 1970 to 1976.' *Demography* 19 (3) 215-334.

34 Hajnal, J (1947) 'The analysis of birth statistics in the light of the recent international recovery of the birth-rate.' *Population Studies* (1) 137-164.

35 Heiland, F, Prskawetz, A and Sanderson, WC (2009) 'Are individuals' desired family sizes stable? Evidence from West German panel data.' *European Journal of Population* 24: 129-156.

36 Iacovou, M and Tavares, LP (2010) 'Yearning, learning and conceding: (some of) the reasons people change their childbearing intentions'. Institute for Social and Economic Research Working Paper 2010-22. University of Essex. Available at: www.iser.essex.ac.uk.

37 Kneale, D and Joshi, H (2008) 'Postponement and childlessness: evidence from two British cohorts.' *Demographic Research* 19: 1935-1968.

38 Schoen, R, Astone, NM, Kim, Y J, Nathanson, CA and Fields, J M (1999) 'Do fertility intentions affect fertility behaviour?' *Journal of Marriage and the Family* 61 (3) 790-799.

39 Testa, R and Toulemon, L (2006) '*Family formation in France: Individual preferences and subsequent outcomes.*' Vienna Yearbook of Population Research: 41-75.

40 Morgan, SP and Rackin, H (2010) 'The correspondence between fertility intentions and behavior in the United States.' *Population and Development Review* 36 (1) 91-118.

41 Testa, R (2007) '*Childbearing preferences and family issues in Europe: evidence from Eurobarometer 2006 survey*.' Vienna Yearbook of Population: 357-379.

National and international graduate migration flows

Irene Mosca *Trinity College, Dublin*

Robert E. Wright *University of Strathclyde, Glasgow*

Abstract

This article examines the nature of national and international graduate migration flows in the UK. Migration equations are estimated with microdata from a matched dataset of *Students and Destinations of Leavers from Higher Education,* information collected by the *Higher Education Statistical Agency*. The probability of migrating is related to a set of observable characteristics using multinomial logit regression. The analysis suggests that migration is a selective process with graduates with certain characteristics having considerably higher probabilities of migrating, both to other regions of the UK and abroad.

Contents

Introduction

One of the key outputs of the higher education sector is the production of skilled labour. It is well-known that, on average, the employment rates and earnings of graduates are considerably above those of non-graduates, suggesting that employers to a certain extent value the skills being generated by the UK higher education sector. It is equally well-known that there is a tendency for graduates to study in and stay after graduation in the region where they studied. However, there is a considerable amount of movement of graduates between different regions of the UK, for example between England, Northern Ireland, Scotland and Wales. Likewise, there is a considerable amount of movement abroad. The main purpose of this article is to quantity the extent of this movement. In addition, an attempt is made to explore empirically the determinants of graduate migration flows.

Data

The analysis is based on microdata collected by *Higher Education Statistical Agency* (HESA)[1]. More specifically, information is merged from two data-sets for five graduation cohorts of higher education institutions (HEI) students, covering the academic years 2002/03 to 2006/07.

The first dataset is the *Students in Higher Education Institutions*[2]. This primarily consists of information provided by the HEI at which the individual studied. As is discussed in more detail below, variables include subject of study, level of study, class of qualification, mode of study, age, gender and place of domicile.

The second dataset is the *Destinations of Leavers from Higher Education Institutions* (DLHE)[3]. The data are collected through a questionnaire, administered approximately six months after the student has graduated. Detailed information about employment and further study is collected.

In this merged dataset, there are three postcodes of interest. The *first* is the postcode corresponding with the individual's so-called 'place of domicile'. This is the postcode of the student's permanent or home address prior to entry to the programme of study. Although imperfect, for the vast majority of graduates this will also be the place where they completed at least some of their secondary schooling. The *second* postcode is 'place of study'. This is simply the address of the HEI attended. The *third* is the postcode that corresponds to their 'place of employment six months after graduation'. Subject to data limitations discussed below, with these three postcodes it is possible to identify if an individual has moved from their place of domicile to their place of study and from their place of study to their place of employment. For those in employment six months after graduation it is possible to calculate migration rates once the level of geographic aggregation has been decided.

The *Destinations of Leavers* survey also interviews graduates who have moved abroad. Therefore, with these data it is not only possible to identify graduates who have moved to other parts of the UK ('national movers'), but also graduates who have emigrated abroad ('international movers'). For the purpose of this article, the level of geographic aggregation for national movers is four countries of the United Kingdom: England, Northern Ireland, Scotland and Wales. Such a division makes considerable sense when it is remembered that Northern Ireland, Scotland and Wales have elected devolved administrations whose responsibilities include matters related to all levels of education. For some analysis, England has been further disaggregated into the nine standard

'NUTS1' regions[4]: East, East Midlands, London, North East, North West, South East, South West, West Midlands and Yorkshire and Humber[5].

There are a series of problems associated with using differing postcodes to proxy migration. Essentially, all these problems manifest measurement error. It is important to note that the survey did not collect any information on the graduate's home address (such as the postcode) at the time of graduation. The only information provided is the postcode of the student's permanent or home address prior to entry to the programme of study. Another problem is that given that 'place of employment' is measured six months after graduation, short-term repeat migration will be missed. For example, a graduate might move from their place of study three months after graduation and then move back to their place of study two months later. This individual would be miss-classified as a 'stayer'. Likewise, for distance learning students, such as those studying at the Open University, allocating place of study would almost certainly miss-classify them as 'movers'. In our analysis, all distance-learning students are assumed to be 'stayers'. There is also a problem dealing with HEIs that have multiple campuses since the data usually only report the name of the institution, with the researcher having to map the postcode. Although it is possible for some cases to identify the geographic location of the campuses, this weakness with the data generates some measurement error. It is also clear that many students commute. For example, it is believed that a sizeable number of students who study at HEIs in London, commute on a regular basis from regions outside London. Commuting is likely to be even more prevalent for students studying on a part-time basis. This is a potential further source of measurement error since for some of these students' place of domicile will not be the same as place of study given postcode information. As a consequence, they would be incorrectly classified as 'movers'.

The *DLHE* survey is a sample of graduates six months after graduation. Since it is a sample, there is always a concern about its overall representativeness. HESA claims a response rate above 75 per cent. They also state that the data are representative of the graduate population. It is hard to substantiate this claim. However, we are not aware of any hard empirical evidence to the contrary. Likewise, our discussion with other researchers using the data does not support the view that the data are non-representative. Nevertheless, graduates who have been more successful in finding appropriate employment, may exhibit a higher response rate. For example, those employed in what are termed 'non-graduate jobs' might have a lower response rate. Along similar lines, those who were less successful might be more reluctant to report details relating to their employer (like postcode). It may also be the case that individuals report the postcode of their firm's head office rather than the postcode of their actual place of work. Finally, there is the problem of those who work remotely, who through the use of information technology and the internet 'work' geographically away from their employer. Again, it is difficulty to establish the seriousness of these problems. However, only a negligible percentage of graduates in employment six months after graduation included in the database did not report all three postcode of interest to us. It is worth noting that the majority of these issues would be less problematic, and could be dealt with in a more systematic way, if the survey collected the postcode of where a graduate actually lives, in addition to the postcode of their place of employment.

The above discussion has highlighted some of the problems using postcode information to identify migration patterns. In this respect it is clear that the data have some important limitations. However, it should be kept in mind that the seriousness of these problems is likely to become more important, when the level of geographic disaggregation is greater. Since the primary focus of the

analysis included in this article is on the movements of employed graduates within and between the four countries of the UK and abroad, we believe that our findings are relevant.

It is important to stress that HESA does not compile similar data for international students, so all the estimates reported in this article refer to UK-domiciled students. In addition, all estimates are reported separately for 'undergraduate graduates' (including individuals being awarded qualifications below degree-level) and 'postgraduate graduates'. Although it is common to pool these two groups together, our analysis suggests that they are quite different. Formal statistical tests (not reported here) indicate that they should be treated as distinct populations, particularly in regression analysis.

Findings

Although this article is primarily concerned with migration after graduation, **Table 1** is a cross-tabulation of country of domicile by country of study for the five graduate cohorts pooled together. If all graduates studied in their country of domicile, then the diagonal cells in this matrix would each be 100 per cent. Although the majority of graduates study in their country of domicile, there is a considerable amount of movement from country of domicile to country of study. For example, of Wales-domiciled undergraduate graduates 33.5 per cent studied in England. Likewise, of Northern Ireland-domiciled postgraduate graduates 20.5 per cent studied in England.

Table 1 Country of domicile by country of study, HEI graduate cohorts, 2002/03 – 2006/07

(a) Undergraduate graduates, per cent

		Country of study			
		England	Northern Ireland	Scotland	Wales
Country of domicile	England	95.4	0.02	1.4	3.2
	Northern Ireland	13.9	75.9	9.6	0.6
	Scotland	6.8	<0.1	93.0	0.2
	Wales	33.5	<0.1	0.6	65.9

Number of observations = 1,159,324

(b) Postgraduate graduates, per cent

		Country of study			
		England	Northern Ireland	Scotland	Wales
Country of domicile	England	96.8	0.2	1.3	1.8
	Northern Ireland	20.5	73.9	4.5	1.1
	Scotland	11.7	0.3	87.4	0.5
	Wales	26.8	0.2	0.8	72.2

Number of observations = 351,547

Source: Authors calculations with HESA data (see text)

Table 2 shows the estimated 'stayer', 'national mover' and 'international mover' rates for both undergraduate and postgraduate graduates for each graduation cohort and for the five cohorts pooled together. The data suggest that for those in employment six months after graduation, the

majority are stayers. The pooled estimates indicate that 92.4 per cent of undergraduate graduates and 92.2 per cent of postgraduate graduates were employed in the same country as they studied. Likewise, 7.6 per cent of undergraduate graduates and 7.7 per cent (of postgraduate graduates had moved. As the table shows, the national mover rate is about twice as large as the international mover rate. It is interesting to note that there is no clear trend from year to year.

Table 2 Stayer, national mover and international mover rates, HEI graduate cohorts, 2002/03 – 2006/07

(a) Undergraduate graduates, per cent

Cohort	Stayer	National Mover	International Mover
2002/2003	92.3	5.2	2.6
2003/2004	92.4	5.3	2.3
2004/2005	92.1	5.5	2.4
2005/2006	92.6	4.9	2.5
2006/2007	92.7	5.0	2.4
All years	92.4	5.2	2.4

(b) Postgraduate graduates, per cent

Cohort	Stayer	National Mover	International Mover
2002/2003	92.1	5.3	2.6
2003/2004	92.3	5.1	2.5
2004/2005	91.9	5.4	2.7
2005/2006	92.4	4.9	2.7
2006/2007	92.4	5.0	2.6
All years	92.2	5.1	2.6

Source: Authors calculations with HESA data (see text)

Table 3 is a cross tabulation of country of study by country of employment six months after graduation. Similar to Table 1, if all graduates were employed in the country in which they studied, the diagonal cells in this matrix would be 100 per cent for each of the four countries and '0 per cent' for the 'Abroad' cell. This is clearly not the case. Again, there is a considerable amount of regional variation. More specifically, graduates of English HEIs have the highest stayer rates. 95.7 per cent of undergraduate graduates and 94.9 per cent of postgraduate graduates are employed in England. The lowest stayer rates are for graduates of Welsh HEIs. Only 61.3 per cent of undergraduate graduates and 64.1 per cent of postgraduate graduates are employed in Wales. As the table shows, about a third of Welch HEI graduates are employed in England. Graduates in Northern Ireland have the highest international mover rates, with 3.9 per cent of undergraduate graduates and 3.4 per cent of postgraduate graduates employed outside the UK.

Table 3 Country of study by country of employment six months after graduation, HEI graduate cohorts, 2002/03 – 2006/07

(a) Undergraduate graduates, per cent

		Country of employment six months after graduation				
		England	Northern Ireland	Scotland	Wales	Abroad
Country of study	England	95.7	0.3	0.6	1.2	2.3
	Northern Ireland	3.4	91.8	0.8	0.1	3.9
	Scotland	11.5	1.5	83.5	0.3	3.3
	Wales	35.7	0.2	0.4	61.3	2.4

Number of observations = 837,279

(b) Postgraduate graduates, per cent

		Country of employment six months after graduation				
		England	Northern Ireland	Scotland	Wales	Abroad
Country of study	England	94.9	0.4	1.0	1.0	2.6
	Northern Ireland	6.6	88.3	1.4	0.3	3.4
	Scotland	11.0	0.7	85.4	0.3	2.6
	Wales	32.5	0.5	0.9	64.1	2.0

Number of observations = 306,924

Source: Authors calculations with HESA data (see text)

Table 4 examines the relationship between country of study and country of employment in more detail. This table shows the distribution of employed graduates broken down further using English regions[6]. It is interesting to note that the share of graduates from Northern Ireland, Scotland and Wales moving to London is not excessively large. The highest rate is 5.5 per cent for postgraduate graduates who studied in Wales. The lowest rate is 0.4 per cent for undergraduate graduates who studied in Northern Ireland. However, when English regions are considered, there is considerable variation. For undergraduate graduates the lowest stayer rate is 41.8 per cent for graduates of HEIs in the South East, with 25.7 per cent of the total moving to London. For undergraduate graduates the highest stayer rate is 71.3 per cent for graduates of London-based HEIs.

The ranking is somewhat different for postgraduate graduates, where the lowest stayer rate is 47.8 per cent for graduates of East Midlands HEIs, with 8.1 per cent of the total moving to London. The highest stayer rate is 71.1 per cent for graduates of North West HEIs. This is slightly higher than the 70.6 per cent for graduates of London HEIs.

Table 4 Distribution of employed graduates six months after graduation, HEI graduate cohorts, 2002/03 – 2006/07

(a) Undergraduate graduates, per cent	Stayed	London	Rest of England	Rest of UK	Abroad
Place of Study					
England	93.5	-	-	2.1	2.5
Northern Ireland	92.9	0.4	1.7	0.9	4.0
Scotland	83.7	3.9	7.0	1.8	3.6
Wales	62.3	4.4	30.2	0.6	2.5
South East	41.8	25.7	28.1	1.9	2.4
East Midlands	42.9	11.2	41.9	1.7	2.3
West Midlands	52.5	11.9	31.4	2.2	2.0
Yorkshire and Humber	54.7	7.8	33.1	1.5	2.9
South West	55.2	13.9	23.8	3.7	3.4
East	58.9	17.9	19.0	1.3	3.0
North East	59.9	8.9	25.3	3.0	3.0
North West	68.1	5.8	20.4	3.5	2.2
London	71.3	-	25.6	0.9	2.2

Number of observations=812,433

(b) Postgraduate graduates, per cent	Stayed	London	Rest of England	Rest of UK	Abroad
Place of Study					
England	94.6	-	-	2.5	2.9
Northern Ireland	90.0	1.1	3.8	1.7	3.5
Scotland	85.5	3.2	7.5	1.0	2.9
Wales	65.1	5.5	26.0	1.4	2.1
East Midlands	47.8	8.1	39.5	2.6	2.1
South East	52.7	22.5	20.0	2.1	2.7
East	55.3	15.9	21.3	2.2	5.2
West Midlands	56.0	9.3	29.9	3.0	1.8
Yorkshire and Humber	61.4	5.4	28.5	2.2	2.5
South West	62.5	10.2	20.3	3.7	3.3
North East	67.6	5.8	20.4	3.6	2.5
London	70.6	-	24.7	1.2	3.5
North West	71.1	4.7	17.5	4.5	2.2

Number of observations=298,136

Source: Authors calculations with HESA data (see text)

Table 5 reports on a further source of variation between country of study and country of employment. More specifically this table shows the stayer, mover and international mover rates broken down by place of study and place of domicile. Basically, the rates are calculated separately for graduates who studied in their country of domicile (for example, England-domiciled students studying in England) and for graduates who studied in a country different to their county of domicile (for example, Northern Ireland, Scotland and Wales-domiciled graduates who studied in England). The latter group is 'Rest of the UK'. What is immediately clear is that there are large differences in rates between 'own-domiciled' and Rest of the UK-domiciled graduates. In all cases, the stayer rate is considerably lower for Rest of the UK-domiciled graduates. For example, for undergraduate graduates who studied in Scotland, the stayer rate for Scotland-domiciled students is 92.0 per cent compared to 37.4 per cent for Rest of the UK -domiciled graduates. In addition, for undergraduate graduates it is always the case that the international mover rate for Rest of the UK-domiciled graduates is higher than for own-domiciled graduates. For postgraduate graduates this is also the case for graduates of English, Welsh and Scottish HEIs, but not for graduates of Northern Irish HEIs.

Table 5 Stayer, national mover and international mover rates by place of study and place of domicile, HEI graduate cohorts, 2002/03 – 2006/07

(a) Undergraduate graduates, per cent

Place of Study	**England**		**Northern Ireland**		**Scotland**		**Wales**	
Place of Domicile	England	Rest of UK	Northern Ireland	Rest of UK	Scotland	Rest of UK	Wales	Rest of UK
Stayer	97.1	53.9	92.0	37.4	90.4	24.3	92.0	50.3
National Mover	0.7	41.9	5.6	54.5	8.6	71.6	4.0	43.9
International Mover	2.2	4.3	2.4	8.1	1.1	4.1	3.9	5.8

(b) Postgraduate Graduates, per cent

Place of Study	**England**		**Northern Ireland**		**Scotland**		**Wales**	
Place of Domicile	England	Rest of UK	Northern Ireland	Rest of UK	Scotland	Rest of UK	Wales	Rest of UK
Stayer	96.6	54.1	93.4	29.5	83.9	17.5	93.3	7.1
National Mover	0.9	40.7	4.6	63.7	14.8	79.0	3.2	90.4
International Mover	2.5	5.3	2.0	6.8	1.4	3.5	3.5	2.5

Source: Authors calculations with HESA data (see text)

There also appears to be a relationship between ‘place of domicile’ and ‘place of employment’. For example, around 13.3 per cent of those who studied in Scotland, Wales or Northern Ireland are England-domiciled graduates who returned to England to work. Of this total, 68.3 per cent returned to the same region of domicile (that is, they returned ‘home’). Put differently, 2.9 per cent of England-domiciled students who moved to Scotland, Wales or Northern Ireland to study returned to England to work. 68.3 per cent of these returned to the same (NUTS1) region of domicile.

Table 6 shows the country distribution of international movers. The European Union is the main destination region. Of those who had moved abroad, 44.1 per cent of undergraduate graduates and 35.6 per cent of postgraduate graduates had moved to the European Union, with France, Germany, Ireland, Italy and Spain being the main destination countries. The most popular destination country for undergraduate graduates is France, accounting for 16.9 per cent of the total, followed by the United States at 9.2 per cent. The most popular destination country for postgraduate graduates is the United States for 14.5 per cent of the total. Somewhat surprisingly, Ireland is the second most popular destination country at 7.1 per cent. At first glance, it may appear 'surprising' that Ireland is the second most popular destination country for postgraduate graduates. The data suggest that there is a considerable amount of movement between Northern Ireland and the Republic of Ireland. In our data period there were 9,019 post-graduates studying in Northern Ireland. Of these, 311 moved abroad to work, which is 3.4 per cent of the post-graduates who studied in Northern Ireland. Of all those moving abroad to work, 63 per cent (196) moved to Ireland.

Table 6 Country distribution of international movers, HEI graduate cohorts, 2002/03 – 2006/07

Country of origin	Undergraduates per cent	Country of origin	Postgraduates per cent
European Union	**44.1**	**European Union**	**35.6**
France	16.9	Ireland	7.1
Ireland	7.0	Germany	5.7
Spain	5.5	France	4.9
Germany	3.8	Spain	3.3
Italy	2.9	Belgium	2.3
Austria	1.9	Holland	2.2
Holland	1.1	Italy	2.2
Belgium	0.9	Greece	1.9
Greece	0.6	Sweden	0.9
Poland	0.5	Cyprus	0.9
Cyprus	0.4	Austria	0.7
Sweden	0.4	Poland	0.5
Czech Rep.	0.4	Finland	0.5
Denmark	0.3	Denmark	0.4
Portugal	0.3	Czech Rep.	0.4
Finland	0.2	Portugal	0.4
Luxembourg	0.2	Luxembourg	0.3
Romania	0.1	Romania	0.2
Hungary	0.1	Hungary	0.2
Malta	0.1	Malta	0.1
Slovakia	0.1	Slovakia	0.1
Bulgaria	0.1	Bulgaria	0.1
Latvia	0.0	Latvia	0.0
Estonia	0.0	Lithuania	0.0
Lithuania	0.0	Estonia	0.0
Slovenia	0.0	Slovenia	0.0
United States	9.2	**United States**	14.5
Japan	6.4	**Africa**	8.4
Canada	5.1	**Australia**	5.2
Africa	5.0	**Canada**	4.0
Australia	4.5	**Japan**	3.1
China	3.9	**Switzerland**	2.9
Latin America	2.9	**China**	2.8
Switzerland	1.9	**New Zealand**	2.5
New Zealand	1.8	**Latin America**	2.3
Hong Kong	1.6	**Hong Kong**	1.3
India	1.3	**India**	0.9
Rest of the world	*12.3*	***Rest of the world***	*16.5*

Source: Authors calculations with HESA data (see text)

Regression analysis

In this section a multinomial logit regression model is used to examine the possible determinants of graduate migration flows. This model is non-linear which implies that its interpretation is less straightforward than for linear regression models. Essentially, it conveniently summarises how the probability of the outcome of interest is related to a set of explanatory variables. In our application the outcome variable takes on three possible values:

(1) Stayer;
(2) National mover; and
(3) International mover, with the reference or baseline category being stayer.

Therefore the estimated effects are relative to this group[7]. In keeping with the descriptive analysis presented above, the model is estimated separately for undergraduate graduates and postgraduate graduates.

The explanatory variables are summarised in **Table 7**. All the variables used in the analysis are categorical, so the table gives the descriptive statistics as percentages. In our view, the use of categorical variables makes the interpretation of the results easier. Although this list of included variables is not complete, the selection does represent factors that others have found to be correlated with migration decisions[8]. The variables considered are:

- Sex
- Mode of study
- Disability status
- Ethnicity
- Class of qualification
- Subject studied
- Type of institution
- Age at graduation
- Moved to study
- Country of domicile
- County of study
- Cohort

Table7 also shows the categories that were chosen as the excluded categories.

Most of these variables are self-explanatory, but several require further explanation. ‘Class of qualification’ was not available for postgraduate graduates. For the variables that had missing data, instead of removing them from the sample, variables representing missing information were created and these were included.

Although it is difficult to interpret the ‘effects’ of these variables, we believe that they help to reduce selection bias resulting from the exclusion of cases with missing information. The categories used for ‘subject studied’ were arrived at after some experimentation.

Table 7 Descriptive statistics, variables included in regression analysis, HEI graduate cohorts, 2002/03 – 2006/07

Variable		Undergraduate graduates per cent	Postgraduate graduates per cent
	Stayer	92.4	92.2
	National mover	5.2	5.1
	International mover	2.4	2.6
Sex	Male	40.0	38.8
	Female	60.0	61.2
Mode of study	Studied full-time	85.0	57.9
	Studied part-time	15.0	42.1
Disability status	Not disabled (excluded)	90.8	91.5
	Disabled	6.8	4.5
	Disabled missing	2.4	3.9
Ethnicity	White (excluded)	84.0	81.1
	Not white	12.3	9.7
	Ethnicity missing	3.7	9.2
Class of qualification	1st class	9.1	--
	2.1 class	39.2	--
	2.2 class (excluded)	25.3	--
	3rd class/Pass/other	12.0	--
	Class missing	14.4	--
Subject of study	Science	45.1	31.7
	Science-led	3.0	0.7
	Social science	22.1	27.3
	Social science-led	2.4	0.3
	Interdisciplinary	1.7	1.3
	Arts/Humanities (excluded)	25.4	38.8
	Arts/Humanities-led	0.4	0.1
	Subject missing	0.1	--
Type of institution	Russell group university	22.3	25.1
	Post-1992 university	39.8	31.2
	Old university (excluded)	25.0	32.9
	Specialist HEI	13.0	10.8
Age at graduation	Age < 24 (excluded)	70.2	23.8
	Age 25-29	10.1	27.6
	Age 30+	19.6	48.7
Moved to study	Yes	50.0	39.3
	No	50.0	60.7
Country of domicile	England (excluded)	82.9	81.7
	Scotland	8.4	9.6
	Wales	4.9	5.0
	Northern Ireland	3.8	3.7
Country of study	England (excluded)	81.9	82.3
	Scotland	9.3	9.6
	Wales	5.8	5.2
	Northern Ireland	3.0	2.9
Cohort	2002/03 (excluded)	19.7	19.0
	2003/04	20.5	19.4
	2004/05	18.3	18.7
	2005/06	20.8	21.2
	2006/07	20.7	21.6

'Science-led', 'Social-science-led' and 'Arts/Humanities-led' refer to joint and mixed qualifications with subjects from these fields dominant.

HEIs are divided into four groups:

- 'Russell' universities belong to a collaboration of twenty leading UK universities that receive around two-thirds of research grant funding in the UK
- 'Old' universities were already classified as universities before 1992 but do not belong to the Russell Group
- 'Post-1992' universities were classified as polytechnics until 1992
- 'Specialist HEIs' include those institutions where subjects including music, dance, drama or art are taught.

The variable 'Moved to study' captures whether the individual moved region to study. In order to construct this variable, England was divided into the nine NUTS1 regions (as discussed above), Scotland was divided into seven regions based on a council area aggregations[9], Wales was divided into three regions (South-, Mid- and North Wales); and Northern Ireland divided into two regions (Belfast and not-Belfast). These regions were constructed in such a way to insure that all regions have HEIs in them. Based on this classification, a graduate was classified as 'moved to study' if their region of domicile was not the same as region of study. It is quite well known in the migration literature that an individual who has moved in the past has a considerably higher probability of moving in the future. Although this variable is crudely measured, it is an attempt to capture this important form of 'path dependence'.

The estimates of the multinomial logit regressions models are summarised in **Table 8**. Because of the very large sample sizes, almost all of the coefficients are statistically significant at conventional threshold levels. For both equations, the pseudo-R^2 values are above 20 per cent, which implies a very good fit remembering that these models are estimated with micro-data. In fact, this is a high value given that the equations were estimated with individual-level data.

Turning first to undergraduates graduates, men compared to women have a higher probability of migrating. That is, men have a higher probability of being both national and international movers, although the effect is most pronounced for international movers. Graduates who studied full-time compared to those who studied part-time have a lower probability of being a national mover but have a higher probability of being an international mover. However, this finding must be viewed with some caution since those studying part-time likely have a higher probability of commuting (as discussed above). Graduates with a disability have a higher probability of migrating and the effect is similar for both types of moves. Being of non-white ethnicity is associated with a lower probability of migrating.

There is a clear gradient with respect to the class of qualification obtained. The higher the class of qualification obtained, the higher the probability of migrating, with the effect being larger on the probability of being an international mover compared to being a national mover.

The results for subject of study are more mixed. Science qualifications (compared to arts and humanities qualifications) are associated with a higher probability of being a national mover but a lower probability of being an international mover. The effect is similar for science-led qualifications

but less pronounced. Interdisciplinary qualifications are associated with a higher probability of migrating but the effect is largest on the probability of being a national mover. Social science qualifications are associated with a lower probability of being an international mover. However, the opposite is the case for social science-led qualifications; this area of study is associated with a higher probability of moving both nationally and internationally. There is little difference between arts and humanities-led qualifications and arts and humanities qualifications. Those who graduated from a 'specialist HEI' (such as an art or music college) have a higher probability of being a national mover and a lower probability of being an international mover. Compared to being a graduate of 'old universities', graduates of Russell Group universities have a higher probability of migrating while graduates of 'Post-1992 universities' have a lower probability of migrating. These effects are much stronger in the opposite directions on the probability of being an international mover. The results suggest that the probability of migrating declines sharply after the age of 30.

Table 8 Multinominal regression results of the probability of migrating, HEI graduate cohorts, 2002/03 – 2006/07

	Undergraduate graduates				Postgraduate graduates			
Variable	National mover	Coeff. in SE*)	Internatio-nal mover	Coeff. in SE*)	National mover	Coeff. in SE*)	Internatio-nal mover	Coeff. in SE*)
Male	0.125	10.7	0.252	17.2	0.169	9.0	0.492	21.0
Studied full-time	-0.091	3.6	0.199	5.0	-0.214	8.9	0.725	23.4
Disabled	0.114	5.2	0.125	4.6	0.023	0.5	-0.094	1.7
Disabled missing	0.227	4.9	0.125	2.1	0.365	7.9	0.300	5.7
Ethnicity Non-white	-0.396	15.1	-0.354	12.6	-0.171	4.2	0.182	4.9
Ethnicity missing	-0.028	0.9	0.155	3.9	0.232	7.7	0.232	6.4
1st class	0.151	7.1	0.400	15.9	--		--	
2.1 class	0.084	5.8	0.246	13.4	--		--	
3rd class/Pass/other	-0.342	16.3	-0.379	12.0	--		--	
Class missing	-0.490	19.2	-0.594	14.6	--		--	
Science	0.107	7.3	-0.614	33.5	0.694	30.8	0.358	12.2
Science-led	0.084	2.3	-0.376	8.8	0.894	8.8	0.298	2.2
Interdisciplinary	0.397	6.9	0.165	2.6	0.315	2.5	0.551	5.1
Arts/Humanities-led	0.116	1.2	-0.069	0.6	-0.720	1.2	0.233	0.5
Subject missing	-0.810	1.4	1.019	3.0	--		--	
Russell group university	0.066	4.5	0.237	13.4	0.098	4.3	0.128	4.9
Post-1992 university	-0.200	12.5	-0.639	30.9	-0.251	9.2	-0.879	24.8
Specialist HEI	0.071	3.5	-0.687	23.6	0.259	7.7	-0.846	15.7
Age at graduation 25-29	-0.037	1.7	-0.093	3.2	0.099	3.9	0.643	21.0
Age at graduation 30+	-0.242	10.6	-0.763	21.0	-0.039	1.4	0.267	7.7
Moved to study	2.462	125.4	0.531	31.8	2.194	88.2	0.451	19.1
Scotland-domiciled	-0.470	18.2	-0.439	10.3	0.203	5.3	-0.111	1.7
Wales-domiciled	1.076	52.6	-0.249	5.9	0.752	22.3	-0.288	3.7
Northern Ireland-domiciled	2.095	68.3	1.267	25.4	1.742	37.5	0.987	12.7
Studied in Scotland	2.472	108.0	0.899	23.9	1.709	47.4	0.031	0.5
Studied in Wales	3.345	211.4	0.589	16.6	3.135	104.7	0.332	4.4
Studied in Northern Ireland	-1.237	27.4	-0.883	14.7	-0.572	9.3	-0.850	8.9
2003/04	0.021	1.2	-0.113	5.0	0.008	0.3	-0.027	0.7
2004/05	-0.034	1.9	-0.110	4.8	-0.012	0.4	0.020	0.6
2005/06	-0.046	2.5	-0.048	2.2	-0.059	2.0	0.049	1.3
2006/07	-0.055	3.0	-0.081	3.6	-0.057	2.0	0.033	0.9
Constant	-5.68	146.3	-3.72	75,8	-5.49	113.1	-4.91	88.5
Number in analysis	*837,279*				*306,924*			
Log likelihood	-202,01				-79,19			
Pseudo R^2	0.236				0.200			

Note: *) Ratio of coefficient to its standard error.

Graduates who moved regions to study have a higher probability of migrating. However, this effect is much larger on the probability of being a national mover compared to being an international mover. There are some clear differences by country of domicile and country of study. As was highlighted in Table 5, there clearly is an interaction between 'country of domicile' and 'country of study' that is not likely captured by the inclusion of dummy variables for each. It is our view that in order to understand how 'country of domicile' and 'country of study' affect the probability of migrating, country and domicile-specific equations need to be estimated. However, this task is outside the scope of this article. Finally, the results suggest that the probability of migrating has declined slightly for the more recent graduate cohorts in the analysis.

In terms of the direction of the effects of the included variables the findings for postgraduate graduates are surprisingly similar. However, there are some differences worth noting. The probability of postgraduate graduates migrating does not appear to be affected by disability status. Graduates of a non-white ethnicity have a higher probability of being an international mover. More recent cohorts of graduates have a lower probability of being a national mover.

How 'big' are the effects of these variables? One way to attempt to answer this is to use the regression equations to 'predict' the probability of migrating nationally and internationally for hypothetical graduates 'made up' of different combinations of the variables included in the regression equations. The obvious baseline for comparison is the hypothetical graduate who represents the mean values of the variables. For this graduate (Graduate A) the predicted probabilities of migrating are the same as the proportions in each category in the raw data (that is, the actual values, see Table 2). For illustrative purposes this hypothetical graduate is compared to one who is a white, non-disabled male who studied full-time, graduated between the ages of 20 and 24 with a 1st class science qualification from a Russell Group university and has moved to another region in order to study (Graduate B).

Table 9 Predicted migration probabilities, HEI graduate cohorts, 2002/03 – 2006/07

(a) Undergraduate graduates, per cent	**Graduate A**	**Graduate B**	**Absolute difference**	**Percentage difference**
Stayer	92.4	82.8	-9.7	-10.4
National mover	5.2	11.7	6.5	126.2
International mover	2.4	5.6	3.1	130.2

(b) Postgraduate graduates, per cent	**Graduate A**	**Graduate B**	**Absolute difference**	**Percentage difference**
Stayer	92.2	81.5	-10.7	-11.6
National mover	5.1	13.2	8.1	157.4
International mover	2.6	5.3	2.7	101.4

Note:

See text for further details.

Graduate A = mean values of variables

Graduate B = white, non-disabled, male, who studied full-time and graduated between the ages of 20 and 24 with a science qualification from a Russell Group university that he moved to in order to study.

The probabilities associated with these two graduates are given in **Table 9**. For both undergraduate graduates and postgraduate graduates, the probability of migrating is over twice as large for Graduate B compared to Graduate A. For undergraduate graduates, the probability of being a national mover increases from 5.2 per cent to 11.7 per cent and the probability of being an international mover increases from 2.4 per cent to 5.6 per cent. As the table shows, the increase in the propensity to migrate in percentage terms is about the same for both types of moves. For postgraduate graduates, the probability of national movers increases from 5.1 per cent to 13.2 per cent and the probability of being an international mover from 2.6 per cent to 5.3 per cent. Again as the table shows in this comparison the impact is larger in percentage terms on moving nationally compared to moving internationally.

Concluding comments

Data collected by the Higher Education Statistical Agency suggest that a large number of UK-domiciled graduates are working outside the United Kingdom six months after graduating. Of the five graduation cohorts spanning the academic years 2002/03 to 2006/07, about 2.4 per cent of undergraduate graduates and 2.6 per cent of postgraduate graduates were working abroad. There is also movement of graduates around the countries of the UK. Data for the same period suggest that about 2.1 per cent of undergraduate graduates of English HEIs are working in Northern Ireland, Scotland or Wales six months after graduation. The analogous estimates for Northern Ireland, Scotland and Wales are much higher at 4.3 per cent, 13.3 per cent and 36.3 per cent, respectively. It is also interesting to note that there is considerable variation in the proportion of graduates who return to their country of domicile after studying in one of the other countries of the UK. For example, 61.9 per cent of England-domiciled students who studied in Northern Ireland, Scotland or Wales returned to England to work. The analogous estimates for Northern Ireland, Scotland and Wales are 6.0 per cent, 60.2 per cent and 85.7 per cent respectively. The figures are similar for postgraduate graduates. The range of estimates is even wider when regions within England are considered.

Regression analysis indicates that the migration of graduates is a selective process. Migration is correlated with a series of characteristics, some of which capture academic performance, such as class of qualification obtained and age at graduation.

Key Findings

- For UK-domiciled graduates there is a considerable amount of movement between country of domicile and country of study.
- Of those graduates who are employed six months after graduation, the majority are employed in the same country where they studied.
- Graduates of English HEIs have the lowest migration rates while graduates of Welsh HEIs have the highest. Graduates of Northern Irish HEIs have the highest international migration rates.
- London is not the main destination of graduates who move nationally, although it is the single most common destination.
- Migration rates are generally much higher for graduates who did not study in their country of domicile, since many return to their region of domicile after graduation.
- The European Union is the most popular destination region for international movers, with the United States also being important.
- A multinomial regression analysis suggests that migration is a selective process with graduates with certain characteristics having considerably higher probabilities of migrating both to other regions of the UK and abroad. Characteristics that appear to be important include class of degree, subject studied, type of institution attended and age at graduation.

Acknowledgements

Financial support from the Economic and Social Research Council under grant: RES-171-25-0032, 'The Overall Impact of Higher Education Institutions on Regional Economies in the UK' is gratefully acknowledged. We are also grateful for the constructive comments received from two anonymous referees.

References

1 HESA is the official agency for the collection, analysis and dissemination of quantitative information about higher education in the United Kingdom. One of its main objectives is to manage a system of data collection, analysis and dissemination aimed at facilitating research. Further information can be found at: www.hesa.ac.uk.

2 For background information and descriptive cross-tabulations see the following annual publication: Students in Higher Education Institutions, Cheltenham, Higher Education Statistical Agency. Available at: www.hesa.ac.uk/index.php?option=com_pubs&Itemid=122

3 For background information and descriptive cross-tabulations see the annual publication: *Destinations of Leavers from Higher Education Institutions*, Cheltenham, Higher Education Statistical Agency. Available at: www.hesa.ac.uk/index.php?option=com_pubs&Itemid=122

4 Eurostat (2009) 'The NUTS classification.' Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/nuts_classification

5 For more detailed information on how the English regions are defined see :Office for National Statistics, Regional Trends. Available at: www.statistics.gov.uk/statbase/product.asp?vlnk=836

6 The sample sizes are different in Tables 3 and 4 because of missing postcode information. Therefore, it was not possible to allocate a specific NUTS1 region to all graduates working in England. The observations were therefore excluded from the calculations of the estimates presented in Table 4.

7 For a comprehensive treatment of the multinomial logit model see Chapter 24 in Greene, W (2007) *Econometric Analysis*. 6th Edition, London, Pearson Education.

8 See for example the studies of: Faggian, A and McCann, P (2006), 'Human Capital Flows and Regional Knowledge Assets: A Simultaneous Equation Approach', Oxford Economic Papers, vol 52: 475-500; Faggian, A McCann, P and Sheppard, S (2006) 'An Analysis of Ethnic Differences in UK Graduate Migration Behaviour'. *Annuals of Regional Science*, vol 40: 461-471; Faggian, A, McCann, P and Sheppard, S (2007) 'Human Capital, Higher Education and Graduate Migration: An Analysis of Scottish and Welsh Students'. *Urban Studies* vol 44: 2511-2528; Faggian, A, McCann, P and Sheppard, S (2007) 'Some Evidence That Woman Are More Mobile than Men: Gender Differences in UK Graduate Migration Behavior'. *Journal of Regional Science,* vol 47: 517-539; and Faggian, A, Li, Q C and Wright, RE (2009) 'Graduate Migration Flows in Scotland'. *Fraser of Allander Economic Quarterly,* vol 33, no 1: 55-60.

9 Aberdeen City Region (Aberdeen City and Aberdeenshire), Dundee City Region (Dundee, Angus and Perth & Kinross), Edinburgh City Region (East Lothian, City of Edinburgh, Midlothian and West Lothian), Stirling (Stirling, Falkirk and Clackmannanshire), Fife, Glasgow City Region (Glasgow, North Lanarkshire, South Lanarkshire, East Renfrewshire, Renfrewshire, Inverclyde, East Dunbartonshire, West Dunbartonshire, East Ayrshire, North Ayrshire, South Ayrshire) and Highlands, Moray & Islands (Argylle & Bute, Highlands, Moray, Eilean Siar, Orkney, Shetland, Scottish Borders and Dumfries & Galloway).

The demographic characteristics and economic activity patterns of carers over 50: evidence from the English Longitudinal Study of Ageing

Athina Vlachantoni

University of Southampton

Abstract

Studies on informal care provision have often focused on the provision of care for persons with a long term physical or mental ill-health or disability, or problems related to old age. However, the provision of care and support more broadly, for example in the form of childcare for grandchildren, can also impact on various aspects of a carer's life, such as their employment (if under the state retirement age), lifetime earnings and, by extension, pension income in later life. This article uses data from Wave 3 of the English Longitudinal Study of Ageing (ELSA) to explore the demographic characteristics, caring patterns, health status and economic activity patterns of carers aged over 50 in England. The results suggest that the nature of care provision differs across age groups, and that caring can be quite a different experience for older men and women. This article also sheds light on the characteristics of 'round-the-clock' carers, a relatively under-researched group which makes up just over one fifth of all carers aged 50 and over.

Contents

Introduction

The activity of caring, whether for one's family, friends or neighbours, has always been part of the human life course. With rapid population ageing affecting both those who receive and those who provide care, researching patterns of informal care provision assumes increasing policy relevance. The 2001 Census showed that about 10 per cent of the UK's population (or 5,884,450 persons) provided unpaid care to "family members, friends, neighbours or others because of long term physical or mental ill-health or disability or problems relating to old age", with a similar proportion for the English population (9.9 per cent of total population)[1]. From the 2000 General Household Survey (GHS) we know that married or cohabiting adults are more likely to be carers than those who are single, or were previously married, and that over one in four carers spends at least 20 hours per week on their caring responsibilities[2]. The 2000 GHS also showed that 18 per cent of women compared to 14 per cent of men were carers, and studies of informal care provision have often highlighted the impact of such gender differences on women's employment, earnings, and by extension, pension entitlements[3 4].

Studies on informal care provision have often focused on the provision of care towards long term, sick or elderly persons, and often on carers aged between 35-64, who constituted approximately 65 per cent of all informal carers in 2001[1]. However, the provision of care can also take other forms, for example, providing childcare for one's grandchild or supporting one's neighbour in performing household tasks. All forms of care provision can impact on the employment patterns of carers prior to retirement, and on the income and pension income of carers more generally, as well as on the quality of life of the person receiving care. In addition, all forms of care provision have implications for the design of social policy supporting both carers and those receiving care. The English Longitudinal Study of Ageing (ELSA) provides a unique opportunity to explore the provision of care, not only towards long term sick, disabled or elderly persons, but also towards other persons, younger or older, who require different kinds and levels of support and care. With evidence of a projected rise in the older part of the population[5] and a rising demand for informal care[6], understanding the nature of informal care provision can lead to better understanding of its impact on the carers' socio-economic status across the life course.

The ELSA dataset also allows exploration of the nature of caring, particularly by those aged 50 and over, an age group that is relatively under-researched in existing literature[7 8]. We know from the 2001 Census that there were approximately 1.24 million men and 1.56 million women over the age of 50 providing unpaid care to sick/disabled persons and that half of these were concentrated among the 'younger old', that is those aged between 50-59[1 9]. Gender differences again permeate these patterns, with women being more likely to be carers than men, except among older ages, and women being more likely to care for persons outside their household[10 11]. Looking at caring in a broader sense, and not only towards long-term sick, disabled or elderly persons, results from the first wave of the ELSA have shown that the majority (56 per cent) of carers aged 50 and over provided up to 19 hours of care per week, about a quarter provided 50 hours of care or more every week, and just over one fifth of all older carers provide care round-the-clock (168 hours per week)[12]. The focus of this article is on persons aged 50 and over as providers rather than receivers of care, and this serves two purposes. First, against evidence of greater role juggling in mid and late life[13], the analysis sheds light on the combination of care provision among the 'younger old' aged between 50 and the state retirement age, some of whom may still be engaged in paid work.

Secondly, the analysis contributes to a growing body of literature, aimed at understanding and recognising the role and contribution of older people as carers[8 14].

Data and methods

The ELSA is a study of people aged 50 and over and their younger partners, living in private households in England. The sample has been drawn from households which previously responded to the Health Survey for England (HSE) between 1998 and 2006. The HSE is an annual cross-sectional household survey which collects a wide range of health data and biometric measures. Each HSE sample is drawn using a two-stage sampling strategy, which involves a selection based on postcodes selected from the Postcode Address File (PAF) and a random selection of households from a fixed number of addresses covering each postcode sector. As a result, the HSE is nationally representative of private households.

There are two main disadvantages in using ELSA data. Firstly, there is the potential loss of representativeness before the ELSA data are drawn from HSE data, due to non-response to HSE, refusal to be re-contacted and attrition between HSE and ELSA. Secondly, as with many national surveys, individuals living in institutions such as residential and nursing homes, are excluded from the ELSA. However, the ELSA aims to look at the circumstances surrounding the move into an institution, and the refreshment of the sample at each wave and the use of weighting partly address the potential loss of representativeness[15].

The analysis presented in this article focuses primarily on Wave 3 (2006) of the ELSA, which at the time of going to print was the most recent wave including a weighting variable. This dataset has a total of 10,513 respondents, of whom 9,771 provided valid responses. These included 7,535 core members (individuals aged 50 and over), 91 core partners, 312 younger partners (partners of core members who were up to 49 years old), 74 'new' partners (who had joined the household since the HSE interview and were found in previous waves) and 26 new partners (who had joined the household since the HSE interview and were found in Wave 3). The initial ELSA sample was refreshed during Waves 2, 3 and 4, in order to make it representative of the youngest people who had aged since each previous wave. The refreshment sample in Wave 3 contains 1,275 core members, 142 older partners, 295 younger partners and 21 new partners (found in Wave 3)[15].

Table 1 uses data from Waves 2, 3 and 4 of the ELSA to show the proportion of different types of carers between 2004-08. Wave 2 (2004) includes a total of 9,432 respondents, and Wave 4 (Phase 1, 2008) includes 10,860 respondents. Wave 1 has been omitted from this comparative table because carers were identified in a different way in this wave, resulting in a significantly different sample population. The article is based on bivariate and multivariate analysis using SPSS. The definitions of key variables used in the analysis in this article are below.

Definition of carers

In Waves 2, 3 and 4 of the ELSA dataset carers were identified in two steps. First, the total sample was asked the following question:

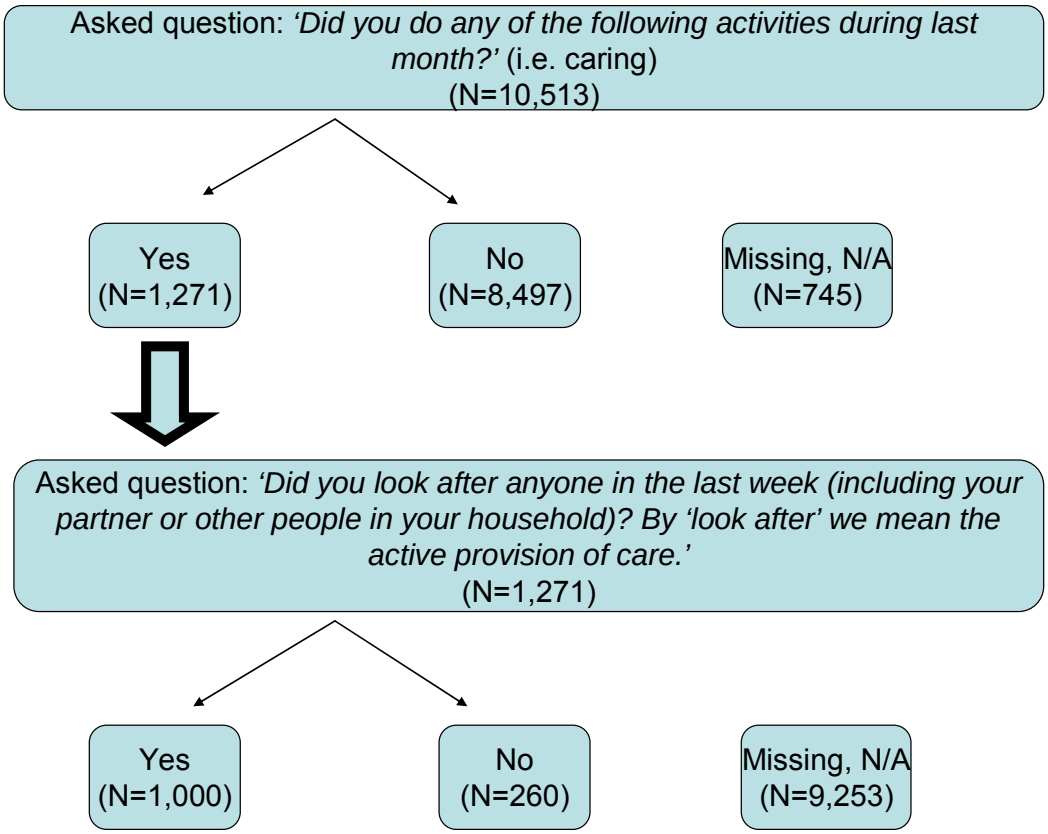
> *'Did you do any of the following activities during last month?'*

Those who included 'caring' in their response to the above question were then asked the following question:

> *'Did you look after anyone in the last week (including your partner or other people in your household)? By 'look after' we mean the active provision of care.'*

The sequence in which the two questions were asked has implications both for the total number of carers identified in the sample and for the types of carers identified. **Figure 1** shows a graphic representation of how carers are identified in Wave 3. The first question yielded a total of 1,271 respondents, who had provided care during the previous month, while the second question reduced the carers' sample to 1,000 persons, who had provided care during both the previous month and during the previous week. Both questions and categories of carers have been taken into account for the analysis in this article. The first part of the article includes a description of the demographic characteristics of carers from both categories (ie. who either cared for another person during the previous week and/or who mentioned caring as part of their activities during the previous month), in order to provide a full picture of caregivers in the dataset. The remainder of the article (Tables 4-12 and Figures 2-4) uses data on those who provided care during the previous week as the unit of analysis. This is because the detailed data on the nature of the activity of caring is only available for this subgroup of carers.

Figure 1 Questions on caring in ELSA, Wave 3



Definition of care intensity

The duration of care provided by older people is measured by the number of hours of care provided per week, ranging from one hour to 168 hours, which represents care provided all the time, or 'round-the-clock'. The majority of carers provided up to 19 hours of care per week, about one fifth provided round-the-clock care and smaller proportions of carers provided care for 20-167 hours per week. The continuous variable has been recoded into four categories of care intensity: 1-19 hours, 20-49 hours, 50-167 hours and round-the-clock care (168 hours per week). This is in order to be consistent with earlier analysis of the ELSA dataset[12], and to be able to explore the group of round-the-clock carers in greater detail.

Relationship to the person receiving care

The ELSA questionnaire asks those who provided care during the previous week to define their relationship to the person receiving care. About one tenth of all carers had looked after more than one person. For this category, co-residence with the person receiving care refers to the main person they looked after, while the total number of hours of care provided includes care for all the persons they looked after. In recognition of the different nature of care provided to different persons, and where larger sample sizes were required, the analysis used a recoded variable which distinguishes between four categories. The first category includes those caring for one's spouse/ partner only, the majority of whom are co-residing with the person they care for. The second category comprises those caring for a parent or a parent-in-law, and in the third category are those caring for a relative, or a friend/neighbour. Finally, the fourth category includes those caring for a child or a grandchild, with the majority of carers in this category caring for a grandchild (**Figure 2**). Although caring for a child in later life may often resemble care for an adult more closely, while care for a grandchild is more likely to be childcare, these two categories of carers are sufficiently different from all other types of carers to be amalgamated for part of the analysis.

Definition of marital status

The 'married' category includes those who are currently married or in a civil partnership (first or subsequent marriage/civil partnership), although the vast majority of this category are married individuals. The dataset does not provide direct information about de facto marital status, which could include a partnership and/or cohabitation outside a legally recognised union.

Definition of health status

Wave 3 of the ELSA includes two questions which are used in this article to explore the health status of carers: self-reported general health status and the report of a limiting longstanding illness. The question on self-reported general health status allowed for five responses (very good, good, fair, bad, very bad), which have been recoded into the following three categories:

Very good/good	= *good*
Fair	= *fair*
Bad/very bad	= *poor*

Definition of economic activity:

The economic activity variable, across all four waves, is derived from the question:

'Which one of these would you say best describes your current situation?'

Economic activity is analysed using five categories (retired/semi-retired; employed/self-employed; unemployed; permanently sick/disabled; and looking after home/family), before distinguishing between those engaged in some kind of 'paid' economic activity (employed, self-employed and unemployed) and those who are not (retired, semi-retired, permanently sick/disabled and looking after home/family).

Results

The results are presented in four parts. The first part explores the demographic characteristics of older carers (for example, age, sex, marital status); the second part explores the nature of the caring activity in greater detail (for example, who they care for, how many hours of care per week, etc); the third part explores the association between the caring activity on one hand, and the carers' health status and their patterns of economic activity on the other; finally, the fourth part investigates the predictors of being a round-the-clock carer for men and women. Part I includes all carers while Parts II-IV focus only on those who report caring for someone in the last week.

Who cares?

Table 1 shows the proportion of those aged 50 and over who provided care between 2004-08, using both definitions used in the ELSA questionnaire to identify carers, ie. those who cared during the previous week and those who cared during the previous month. The table shows a stable proportion of men who provided care during the previous week between 2004-06 (7.8 per cent and 7.9 per cent), and a decline reflected in the 2008 unweighted data (6.7 per cent).

Table 1 Carers and non-carers, by sex, 2004-08

	2004 [a]	2006 [b]	2008 [c]
Men			
Cared last week	7.8	7.9	6.7
Cared last month	2.2	2.0	2.4
Did not provide care	90.0	90.0	90.9
Number in analysis	*4,126*	*4,291*	*4,832*
Women			
Cared last week	13.1	12.3	11.0
Cared last month	3.6	2.9	3.9
Did not provide care	83.3	84.8	85,1
Number in analysis	*5,306*	*5,466*	*5,997*

a: $\chi 2=82.163$, df=2, $p<0.001$; b: $\chi 2=52.545$, df=2, $p<0.001$; c: $\chi 2=82.168$, df=2, $p<0.001$

2004-06: Weighted percentages and unweighted frequencies.

2008: Unweighted percentages and frequencies.

Source: ELSA, Waves 2-4

For men who provided care during the previous month, and for women who provided care during the previous week or month, the 2004-06 data show a small decline, while the 2008 unweighted data show a small increase for men and women who provided care during the previous month. The proportion of the sample who did not provide care has remained constant around 90 per cent for men and around 84 per cent for women between 2004-08. Finally, women are more likely than men to have provided care across the three waves, whether during the previous week or during the previous month.

Patterns of care provision vary for different age groups, but also in different ways for men and women. In 2006 the proportion of all carers in the population aged 50 and over decreased in line with increasing age. **Table 2** shows that just over 15 per cent of those aged 50-59 had provided care during the previous week or month, compared to 13 per cent of those aged 60-69, 12 per cent of those aged 70-79 and almost six per cent of those aged 80 and over. Women were more likely to provide care than men in every age group except among the oldest (those aged 80 and over), where men were more likely to be carers.

Table 2 Care status by age group and sex, 2006

	50-59 years	60-69 years	70-79 years	80 years and over
Men[a]				
Cared last week	7.4	7.8	9.3	7.8
Cared last month	2.9	1.6	1.9	[0.7]
Did not provide care	89.7	90.6	88.8	91.5
Number in analysis	*1,542*	*1,279*	*929*	*453*
Women[b]				
Cared last week	16.6	13.9	10.0	3.2
Cared last month	3.8	3.4	2.1	[1.1]
Did not provide care	79.6	82.7	88.0	95.7
Number in analysis	*1,887*	*1,423*	*1,134*	*680*
All carers[c]				
Cared last week	*12.1*	*10.9*	*9.6*	*4.9*
Cared last month	*3.3*	*2.6*	*1.9*	*1.0*
Did not provide care	*84.6*	*86.6*	*88.5*	*94.1*
Number in analysis	*3,429*	*2,702*	*2,063*	*1,133*

a: $\chi 2$= 12.623, df=6, p<0.010; b: $\chi 2$=109.311 , df=6, p<0.001; c: $\chi 2$=73.371 , df=6, p<0.001

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

The association between being married and providing care was evident in the ELSA dataset, but so were gender differences in this respect. Men who provided care were more likely to be married than women who provided care (85 per cent of all male carers and 77 per cent of all female carers were married (p<0.001)). Women were twice as likely as men to provide care, regardless of their marital status (**Table 3**). The only exceptions were widows and widowers, who were almost equally likely to be carers. The results show that women are more likely than men to provide care during the previous month or week. They are more likely to provide care across different age groups,

except over the age of 80, and are more likely to provide care regardless of their marital status, (except for widows).

Table 3 Care status by marital status and sex, 2006

	Single/never married	Married/civil partnership	Divorced/ seprarated	Widowed
Men[a]				
Cared last week	4.6	9.1	5.2	3.9
Cared last month	[1]	2.1	2.2	[2]
Did not provide care	94.4	88.8	92.6	94.1
Number in analysis	*282*	*3,216*	*412*	*381*
Women[b]				
Cared last week	9.7	16.2	10.2	4.4
Cared last month	[2.1]	3.2	3.7	2.0
Did not provide care	88.2	80.6	86.1	93.6
Number in analysis	*291*	*3,310*	*674*	*1,191*

a: $\chi 2$= 24.012, df=6, p<0.001; b: $\chi 2$= 113.012, df=6, p<0.001

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

These results confirm a strong relationship between caring and gender in later life, which has been found elsewhere. For example, Glaser and Grundy (2002)[11] found that women are more likely to provide care, especially outside the household, among those aged 55-69, while Arber and Ginn (1995)[10] also showed that women are more likely than men to care for someone outside the household. Dahlberg et al (2007)[8] found that men aged 70 and over are more likely to be carers than women and Del Bono et al (2009)[16] attributed men's greater likelihood to provide care at older ages to gender differences in marital status at those ages, that is, a greater likelihood for older men to have a partner to care for. Finally, Young et al (2006)[17] studied couples of which one of the spouses suffered from a LLSI, and found that women were slightly more likely to provide care than men.

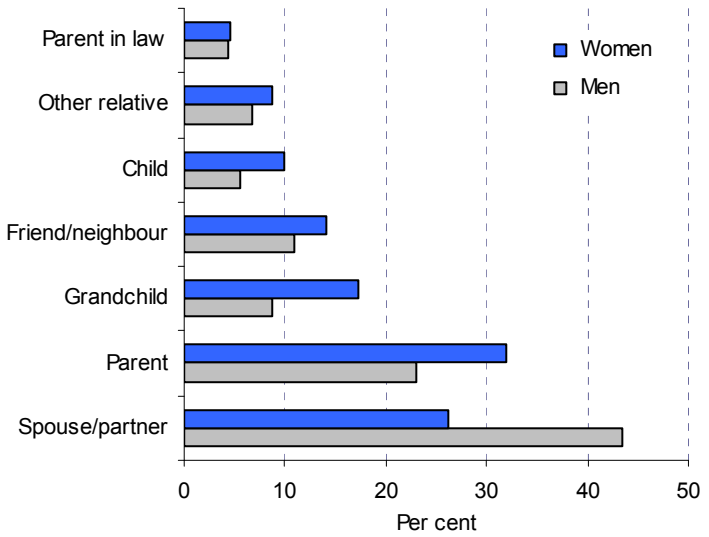
Researching the caring activity in greater detail reveals even more differences in the ways men and women experience care provision in later life. The remainder of the analysis focuses on those who provided care during the previous week.

The nature of the caring activity

This section explores the nature of the caring activity, including the relationship to the person receiving care, the number of hours of care provided per week and the prevalence of co-residence with the person receiving care. **Figure 2** shows that a greater proportion of female carers than male carers looked after parents, grandchildren, friends/neighbours, children, other relatives or parents-in-law, while men were more likely to be caring for their spouse/partner. For example, about one third of women cared for a parent compared to 23 per cent of men, and about 17 per cent of women cared for a grandchild compared to nine per cent of men, but 43.5 per cent of men compared to 26 per cent of women cared for their spouse/partner. In Figure 2 the total proportion of women who looked after different categories of care recipients adds up to more than 100 per

cent, because women are more likely than men to be juggling multiple caring roles. **Figure 3** shows that 72 per cent of women compared to 85 per cent of men cared for one person only, while 16 per cent of women compared to eight per cent cared for two persons, five per cent of women compared to four per cent of men cared for three persons, and three per cent of all female carers looked after four persons. This adds a clear gender dimension to the evidence showing multiple role performance in mid and late life[13], at least in so far as caring over the age of 50 is concerned.

Figure 2 Relationship to the person receiving care, by sex, 2006



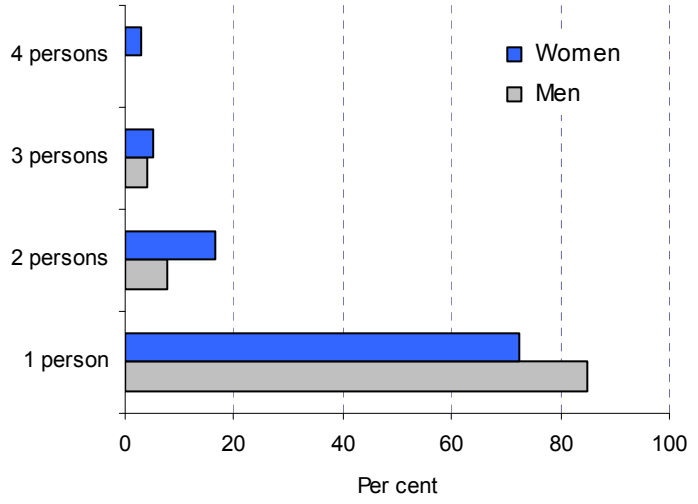
χ2=44.334, df=6, p<0.001.

Weighted percentages. Male carers (N= 334); Female carers (N= 666)

Note: Percentages add up to more than 100 per cent as some people care for more than one individual.

Source: ELSA, Wave 3

Figure 3 Number of persons receiving care, by sex, 2006



Men: χ2=4057.000, df=20, p<0.001. Women: χ2=4604.000, df=42, p<0.001

Source: ELSA, Wave 3

Table 4 suggests that the relationship of the carer with the person they care for changes for different age groups. There are significant differences in the way these changes are manifested among male and female carers. 'Younger old' carers (50-59) were more likely to care for parents/ parents-in-law or for children/grandchildren, while 'older old' carers (80 and over) were more likely to care for their spouse/partner or for other relatives/friends/neighbours. Just over half of all carers aged 50-59 cared for a parent/parent-in-law, and one fifth of this age group cared for a child/ grandchild. The older the carer, the more likely they were to be looking after their spouse/partner, compared to the combined categories of caring for a parent/parent-in-law or a child/grandchild. In the oldest age group almost two thirds of carers aged 80 and over looked after their spouse/ partner, and almost one quarter of carers in this age group looked after another relative/friend/ neighbour. At the same time the changes between different age groups in the relationship with the person receiving care happen in different ways for men and women. Almost one quarter of male carers aged 50-59 looked after their spouse/partner, compared to just over one tenth of female carers in that age group. Women aged 50-59 and 60-69 were more likely than men in these age groups to care for a child/grandchild, while men in all age groups were more likely than women to care for their spouse/partner. When care intensity it taken into account, the differences between male and female carers become even more striking.

Table 4 Relationship to person receiving care by age and sex, 2006

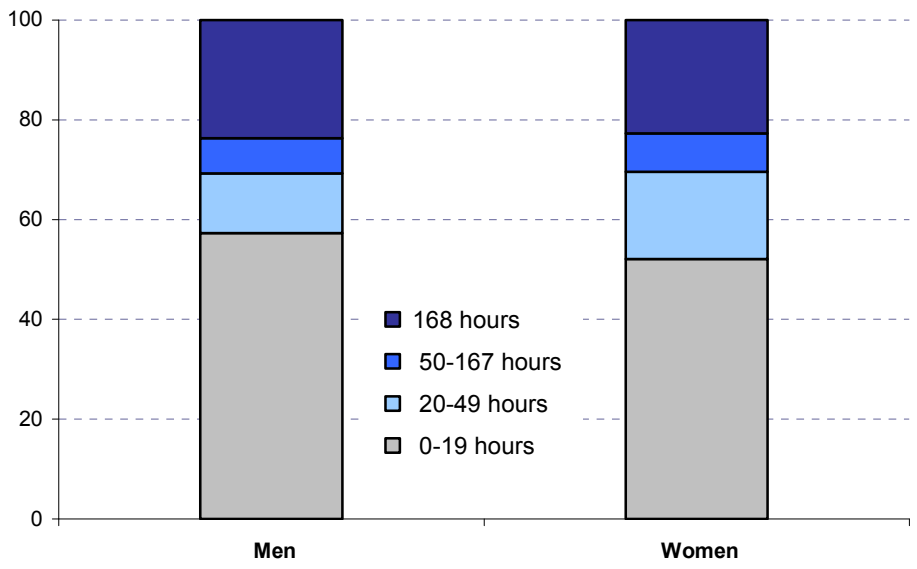
	50-59 years	60-69 years	70-79 years	80 years and over
All carers[a]				
Cared for spouse/partner	14.8	31.4	42.9	61.8
Cared for parent/parent-in-law	50.7	26.8	7.9	[0.1]
Cared for other rel./friend/neighbour	14.6	20.0	33.9	23.6
Cared for child/grandchild	19.9	21.8	15.3	[14.5]
Number in analysis	*406*	*290*	*194*	*53*
Men[b]				
Cared for spouse/partner	24.1	42.7	58.4	65.6
Cared for parent/parent-in-law	45.4	31.3	[3.9]	-
Cared for other rel./friend/neighbour	15.7	12.5	26.0	[21.9]
Cared for child/grandchild	14.8	13.5	11.7	[12.5]
Number in analysis	*112*	*98*	*81*	*31*
Women[c]				
Cared for spouse/partner	10.8	25.5	31.3	56.5
Cared for parent/parent-in-law	53.0	24.5	11.1	-
Cared for other rel./friend/neighbour	14.1	23.9	40.4	[26.1]
Cared for child/grandchild	22.1	26.1	17.2	[17.4]
Number in analysis	*294*	*192*	*113*	*22*

a: $\chi 2$= 175.975, df=9, p<0.001; b: $\chi 2$=61.637, df=9, p<0.001; c: $\chi 2$=111.994, df=9, p<0.001

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

Figure 4 Care intensity by sex, 2006



2006: $\chi 2$=5.193, df=3, p=0.158

Source: ELSA, Wave 3

When considering the intensity of care provided, the ELSA data show that about 56 per cent of all carers provided up to 19 hours of care per week, about 15 per cent provided between 20-49 hours of care, almost eight per cent provided between 50-167 hours, and finally just over one fifth provided round-the-clock care. However, there are certain gender differences as **Figure 4** shows. Women were less likely than men to provide up to 19 hours of care, and more likely than men to provide between 20-49 hours of care. Men and women over the age of 50 were almost as likely to provide between 50-167 hours of care and to be round-the-clock carers.

We already know that for women the 'younger old' are more likely to be carers than the 'older old' (Table 2), but age appears to be also associated with the number of hours of care provided, with the majority of less intense carers being younger old, but more than half of round-the-clock carers are aged 65 and over (**Table 5**). These patterns were broadly similar for men and women, with a couple of exceptions; for example, among those caring between 20-49 hours per week, 55 per cent of men compared to 72 per cent of women were aged 50-64, and among round-the-clock carers, one third of men compared to one half of women were aged 50-64.

Analysing care intensity alongside the relationship with the person receiving care provides further insight in the gendered nature of care giving in older ages. **Table 6** shows that among those caring for up to 19 hours per week, about two fifths of both men and women cared for a parent/parent-in-law, while among those caring for 20-49 hours per week, two fifths of women looked after a child/grandchild and two the same proportion of men looked after their spouse/partner.

Gender becomes a crucial differentiating factor when it comes to round-the-clock care, as although the majority of all round-the-clock carers provided care to their spouse/partner, this was more the case for men than for women (88 per cent of men compared to 57 per cent of women). Female round-the-clock carers were much more likely to look after a child/grandchild (22 per cent).

Table 5 Care intensity by age group, 2006

	50-64 years	65-79 years	80 years and over	*Number in analysis*
All carers[a]				
0-19 hours	65.8	30.4	3.8	*523*
20-49 hours	66.9	27.9	[5.2]	*147*
50-167 hours	54.5	37.9	[7.6]	*74*
168 hours (round-the-clock)	42.1	45.5	12.4	*205*
Men[b]				
0-19 hours	63.3	31.1	5.6	*187*
20-49 hours	55.3	34.2	[10.5]	*39*
50-167 hours	52.2	34.8	[13.0]	*23*
168 hours (round-the-clock)	29.3	49.3	21.3	*74*
Women[c]				
0-19 hours	67.5	29.8	[2.7]	*336*
20-49 hours	72.2	24.7	[3.1]	*108*
50-167 hours	55.8	39.5	[4.7]	*51*
168 hours (round-the-clock)	50.0	42.9	[7.1]	*131*

$\chi 2$= 43.348, df=6, p<0.001; b: $\chi 2$=29.189, df=6, p<0.001; c: $\chi 2$=17.708, df=6, p<0.001

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

Table 6 Relationship to the person receiving care by care intensity and sex, 2006

	0-19 hours	20-49 hours	50-167 hours	168 hours
Men[a]				
Cared for spouse/partner	19.8	38.9	72.7	87.7
Cared for parent/parent-in-law	37.3	[27.8]	[9.1]	[4.3]
Cared for other rel./friend/neighbour	28.8	[11.1]	-	[1.4]
Cared for child/grandchild	14.1	[22.8]	[18.2]	[6.8]
Number in analysis	*187*	*40*	*23*	*73*
Women[b]				
Cared for spouse/partner	6.6	12.5	33.3	57.0
Cared for parent/parent-in-law	41.5	38.5	28.6	15.7
Cared for other rel./friend/neighbour	35.6	9.4	[11.9]	[5.0]
Cared for child/grandchild	16.3	39.6	26.2	22.3
Number in analysis	*360*	*106*	*52*	*129*

a: $\chi 2$= 118.694, df=9, p<0.001; b: $\chi 2$= 194.323, df=9, p<0.001

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

The caring experience, including its intensity, precise tasks as well as the extent to which it is supported by the welfare state, can be different depending on whether the carer lives with the person care is provided for. The ELSA data show that about 40 per cent of carers aged 50 and over lived with the person they cared for, and that the prevalence of co-residence increases in line with a rise in care intensity. For example, less than one fifth of those caring for up to 19 hours lived

with the person they cared for, compared to one third of those caring for 20-49 hours, two thirds of those caring for 50-167 hours and the vast majority of round-the-clock carers.

Co-residence also varied according to the relationship with the person receiving care. **Table 7** shows the proportion of carers who lived with the person receiving care, based on their relationship to that person. Perhaps not surprisingly, almost all of those caring for their spouse/partner lived with them, compared to about one third of those caring for a child/grandchild, one fifth of those caring for apparent/ parent-in-law and about nine per cent of those caring for another relative, friend or neighbour. There were almost no gender differences among carers in this respect, except for women caring for a child/grandchild, who were far more likely than men to live with them (33 per cent compared to 26 per cent).

Table 7 Relationship to person receiving care, by co-residence, 2006

	Lived with them	Lived apart	*Number in analysis*
All carers[a]			
Cared for spouse/partner	99.2	[0.8]	*260*
Cared for parent/parent-in-law	19.3	80.7	*317*
Cared for other rel./friend/neighbour	9.8	90.2	*204*
Cared for child/grandchild	31.3	68.7	*191*
Men[b]			
Cared for spouse/partner	99.2	[0.8]	*137*
Cared for parent/parent-in-law	20.5	79.5	*88*
Cared for other rel./friend/neighbour	[7.5]	92.5	*56*
Cared for child/grandchild	26.2	73.8	*42*
Women[c]			
Cared for spouse/partner	99.2	[0.8]	*123*
Cared for parent/parent-in-law	19.1	80.9	*229*
Cared for other rel./friend/neighbour	10.7	89.3	*148*
Cared for child/grandchild	33.1	66.9	*149*

a: $\chi 2$= 473.266, df=3, p<0.001; b: $\chi 2$=205.374, df=3, p<0.001; c: $\chi 2$=256.196, df=3, p<0.001

Notes: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

Finally, a carer's marital status was also found to be associated with the relationship to the person receiving care, again with certain gender differences (**Table 8**). The vast majority of men and women who were caring for their spouse/partner were married, as were 79 per cent of both men and women who cared for a parent/parent-in-law. Divorce and widowhood appeared to be more prevalent among female carers, with almost a quarter of women caring for a relative/friend/ neighbour or a child/grandchild being divorced and almost one quarter of women caring for a relative/friend/neighbour being widowed.

Moving beyond the characteristics of carers and the nature of the caring activity, the third section of this article turns to the relationship between caring on one hand, and health status and economic activity patterns on the other.

Table 8 Relationship to person receiving care by marital status and sex, 2006

	Single/never married	Married/civil partnership	Divorced/ seprarated	Widowed	*Number in analysis*
Men[a]					
Cared for spouse/partner	-	96.2	[3.0]	[0.8]	*137*
Cared for parent/parent-in-law	12.2	79.3	[6.1]	[2.4]	*88*
Cared for other rel./friend/neighbour	[1.8]	73.2	[10.7]	[14.3]	*60*
Cared for child/grandchild	[2.3]	81.4	[9.3]	[7.0]	*43*
Women[b]					
Cared for spouse/partner	[1.7]	96.6	[0.8]	[0.8]	*123*
Cared for parent/parent-in-law	[4.3]	78.7	13.3	[3.7]	*230*
Cared for other rel./friend/neighbour	10.5	57.3	8.9	23.4	*152*
Cared for child/grandchild	-	75.8	14.5	9.7	*149*

a: $\chi 2$= 46.609, df=9, p<0.001; b: $\chi 2$= 89.670, df=9, p<0.001

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

Health status and economic activity patterns of carers

Health status

The health status of carers is measured using the two variables of self-reported general health (derived as 'good, 'fair' or 'poor') and the reporting of a limiting longstanding illness (LLSI) (yes/no). The ELSA data present a complicated relationship between health status and the provision of care, particularly when care intensity is taken into account (**Table 9**). In the first instance, carers appear more likely than non-carers to report good health, and seem less likely than non-carers to report fair or poor health. This pattern is mirrored in the reporting of a LLSI, although these results are not statistically significant. These results are replicated in Waves 1, 2 and 4, and are broadly similar (but not statistically significant) for men and women. Table 9 also shows an inverse relationship between good health and care intensity, with the proportion of carers reporting good health decreasing as the number of hours of care provided increases. The relationship between care intensity and the report of a LLSI is less straightforward, with a lower proportion of round-the-clock carers reporting a LLSI than those caring for 50-167 hours per week. The generally good or fair health status of carers may be due to the nature of care explored in this dataset, which is of a more general kind and also includes childcare, while the relatively good heath status of round-the-clock carers may be considered a prerequisite for providing care round-the-clock. This category's relatively good health status may also be partly due to the fact that the vast majority of round-the-clock carers care for their spouse/partner and are married, and marriage is associated with a better health status[18].

Table 9 Self-reported health status and report of a LLSI by care intensity, 2006

	Good	Fair	Poor	*Number in analysis*
Caring[a]				
Yes	71.4	24.0	4.6	*1,000*
No	66.3	26.0	7.7	*8,533*
Care intensity[b]				
0-19 hours	77.8	19.1	3.2	*555*
20-49 hours	72.1	22.1	5.9	*149*
50-167 hours	60.6	33.3	[6.1]	*76*
168 hours (round-the-clock)	59.4	34.7	5.9	*208*

Caring[c]	With Limiting Long Standing Illness (LLSI)	No LLSI	*Number in analysis*
Yes	53.8	46.2	*1,000*
No	55.3	44.7	*8,763*
Care intensity[d]			
0-19 hours	50.0	50.0	*555*
20-49 hours	47.1	52.9	*149*
50-167 hours	72.7	27.3	*76*
168 hours (round-the-clock)	60.9	39.1	*208*

a: $\chi 2$=15.129, df=4, p<0.005; b: $\chi 2$=28.355, df=6, p<0.001;

c: $\chi 2$=1.481, df=2, p=0.477; d: $\chi 2$=18.825, df=3, p<0.001

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

Economic activity

The ELSA data show a statistically significant relationship between economic activity and care provision, as well as between economic activity and the intensity of care provision. **Table 10** shows that, among those aged 50-64, carers were less likely to be employed or self-employed than non-carers. However, when we explore economic activity patterns among carers only, and according to the number of hours of care provided, a more complex picture emerges.

Table 10 Economic activity 50-64 year olds by care intensity, 2006

	Employed/Self-employed	Unemployed	Retired/Semi-retired	Permanently sick/disabled	Looking after home/family	Number in analysis
Caring[a]						
Yes	50.8	[1.1]	23.2	5.1	19.8	*581*
No	64.4	1.9	18.5	9.1	6.1	*4,319*
Care intensity[b]						
0-19 hours	62.3	-	18.4	5.8	13.5	*346*
20-49 hours	42.9	[4.4]	31.9	[2.2]	18.7	*97*
50-167 hours	30.6	-	50.0	[2.8]	[16.7]	*40*
168 hours (round-the-clock)	23.3	-	22.1	[8.1]	46.5	*92*

[a]: χ^2=141.291, df=4, p<0.001; [b]: χ^2=101.399, df=12, p<0.001

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

The proportion of carers who were employed or self-employed decreased as the number of hours of care provided rose, however the proportion of those providing up to 19 hours of care per week who were employed or self-employed was almost similar to the equivalent proportion of non-carers (62 per cent compared to 64 per cent). One of the most striking results of this analysis is that almost one quarter of round-the-clock carers aged 50-64 were employed or self-employed, and additional analysis showed that the majority of those (four fifths) were employed.

Table 11 focuses on the population aged 50-64, distinguishing between those who were engaged in some kind of economic activity (employed, self-employed or unemployed) and those who were not (retired, semi-retired, permanently sick/disabled and those looking after home/family). Two thirds of all non-carers aged 50-64 were economically active, compared with just over half of all carers. This is true for both men and women. For example, 73 per cent of male non-carers were economically active, compared to 57 per cent of male carers. Similarly, 59 per cent of female non-carers were economically active, compared to 49 per cent of female carers.

Table 11 Economic activity 50-64 year olds by sex and care intensity, 2006

	Active	Inactive	*Number in analysis*
All - Caring[a]			
Yes	51.8	48.2	*581*
No	66.3	33.7	*4,319*
Care intensity[b]			
0-19 hours	62.5	37.5	*346*
20-49 hours	47.8	52.2	*97*
50-167 hours	30.6	69.4	*40*
168 hours (round-the-clock)	23.5	76.5	*92*
Men – Caring[c]			
Yes	56.7	43.3	*174*
No	73.3	26.7	*2,076*
Care intensity[d]			
0-19 hours	67.3	32.7	*117*
20-49 hours	[42.9]	57.1	*20*
50-167 hours	[41.7]	[58.3]	*11*
168 hours (round-the-clock)	[22.7]	77.3	*24*
Women – Caring[e]			
Yes	49.4	50.6	*407*
No	58.8	41.2	*2,243*
Care intensity[f]			
0-19 hours	59.4	40.6	*229*
20-49 hours	49.3	50.7	*77*
50-167 hours	[29.2]	70.8	*29*
168 hours (round-the-clock)	23.8	76.2	*68*

a: $\chi2=42.661$, df=1, $p<0.001$; b: $\chi2=48.292$, df=3, $p<0.001$;

c: $\chi2=21.531$, df=1, $p<0.001$; d: $\chi2=18.199$, df=3, $p<0.001$;

e: $\chi2=10.638$, df=1, $p<0.005$; f: $\chi2=28.288$, df=3, $p<0.001$

Note: Percentages in brackets are based on cell counts below 10.

Source: ELSA, Wave 3

When the intensity of care provision was taken into account, the prevalence of economic activity decreased as the number of hours of care provided increased. For example, 62 per cent of all carers providing between 0-19 hours of care were active, compared to 48 per cent of those caring for 20-49 hours, 31 per cent of those caring for 50-167 hours and one quarter of the round-the-clock carers. Analysis of the 50-64 year old population by age group (50-54, 55-59 and 60-64), which was not statistically significant and not shown here, showed a decrease in economic activity among both carers and non-carers at key age stages in the life course, namely among women aged 60-64 who enter retirement, and among men aged 60-65 who take early retirement. Certain gender differences are worth noting: Among less intense carers (0-19 hours), 67 per cent of men compared to 59 per cent of women were economically active. Finally, almost half of women caring for 20-49 hours and almost one quarter of female round-the-clock carers were economically active.

What makes a round-the-clock carer?

The most intriguing results in this article refer to the round-the-clock carers, who constitute just over one fifth of all carers in this wave, are more likely to report good health than less intense carers, are roughly evenly distributed among men and women, and one quarter of whom, among those aged 50-64, are engaged in economic activity. To investigate the relative importance of factors affecting one's chances of caring round-the-clock, the final part of the analysis in this article estimated the odds of being a round-the-clock carer for men and women separately (**Table 12**). The findings suggest that care-giving over the age of 50 is quite a different experience for men and women. The model was built in six stages, taking age (by 10 year age groups), marital status, health status, the reporting of a LLSI, economic activity and the relationship to the person cared for into account, and it shows that the predictors of being a round-the-clock carer are quite different for men and women. The logistic regression shows that the relationship with the person receiving care is by far the strongest predictor of being a round-the-clock carer for both men and women, with its inclusion resulting in the largest reduction in the log likelihood ratio. Those caring for their spouse/partner were significantly more likely to be round-the-clock carers than any other group.

When exploring other predictors of being a round-the-clock carer, the gendered nature of care provision over the age of 50 becomes more apparent. For example, age is a stronger predictor for male carers until economic activity is entered in the model, while marital status remains a strong predictor at the variable level for female carers even when economic activity is taken into account. The odds of men aged 70-79 being round-the-clock carers are 4.7 times the odds of men aged 50-59, while the odds of men aged 80 and over being round-the-clock carers are 5 to 7 times the odds of men aged 50-59. This confirms men's increasing likelihood to care in older ages and their greater likelihood to care for their spouse/partner. Age is not a significant predictor for women, perhaps because women are more likely than men to provide care throughout older age and not only in the oldest age group. By contrast, marital status is a much stronger predictor of being a round-the-clock carer among women. For example, the likelihood of being a round-the-clock carer for widowed women was 88 per cent lower than for single, never married women (model 5).

The relationship between health status and round-the-clock care provision remains difficult to interpret from the results. Reporting a LLSI appears to be not significant as a predictor for either men or women, while health status is generally not significant, with one exception. Women who reported 'fair' health were 78 per cent more likely than women reporting 'good' health to be round-

the-clock carers (model 4), but this effect is not significant, once economic activity and relationship to the person receiving care are controlled for.

Finally, economic activity is a strong predictor for both men and women. For women it was, even when the relationship with the person receiving care, is taken into account. The models show an interesting relationship between employment and round-the-clock care. Employed men and women are between 78-82 per cent less likely than retired/semi-retired men and women to be round-the-clock carers (model 5). As we might expect, looking after a home/family made a carer much more likely to be caring round-the-clock, but this effect was greater for men. The odds of men who look after the home/family being round-the-clock carers were eight times those of retired/semi-retired men. The odds of women who look after the home/family of being round-the-clock carers were three times those of retired/semi-retired women. A detailed exploration of the employment patterns of male and female carers, including a breakdown of part- and full-time work, would be necessary to analyse further the relationship between economic activity and intense care provision.

Table 12a **Odds ratios of being a round-the-clock carer, men, 2006**

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Age group						
50-59 (ref)	1.00***	1.00***	1.00***	1.00***	(ns)	(ns)
60-69	2.18	2.18	2.11	1.99		
70-79	5.05***	5.31***	5.12***	4.73***		
80 and over	5.38***	7.40***	7.11***	6.69***		
Marital status						
Single, never married (ref)	-	(ns)	(ns)	(ns)	(ns)	(ns)
Married/civil partnership						0.04**
Divorced/separated						
Widowed						
Health status	-	-	(ns)	(ns)	(ns)	(ns)
LLSI	-	-	-	(ns)	(ns)	(ns)
Economic activity						
Retired/semi-retired (ref)	-	-	-	-	1.00***	(ns)
Employed					0.18***	
Self-employed					0.16	
Permanently sick/disabled					0.96	
Looking after home/family					7.98***	
Relationship to person receiving care						
Spouse/partner (ref)	-	-	-	-	-	1.00***
Parent/parent-in-law						0.02***
Other relative/friend/neighbour						0.02***
Child/grandchild						0.17***
Constant	0.007***	0.005***	0.005***	0.007***	0.012***	21.053
Log Likelihood ratio	719.4	700.7	694.2	687.8	649.9	223.7

* p < 0.010

** p < 0.005

*** p < 0.001

(ns) = not significant.

[a] Note: 'Unemployed' was excluded from the models, as there were no unemployed among the round-the-clock carers.

Source: ELSA, Wave 3

These results confirm the gendered pathways of men and women into care provision in later life. Round-the-clock caring by women is largely determined by who they care for, but also by their marital and economic activity status. Earlier in this article we saw that women provide more care in every age group, and, if round-the-clock carers, in about half of the cases they look after a parent/ parent-in-law or a child/grandchild. For men it is the relationship to the person they care for, as well as their age and economic activity status which determine whether they will become round-the-clock carers. Indeed, for the majority of men, intense care provision is an experience which comes in older ages, and which is predominantly focused on their spouse/ partner.

Table 12b **Odds ratios of being a round-the-clock carer, women, 2006**

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Age group	(ns)	(ns)	(ns)	(ns)	(ns)	(ns)
Marital status						
Single, never married (ref)	-	1.00***	1.00***	1.00***	1.00***	(ns)
Married/civil partnership		1.46	1.53	1.51	1.08	0.15**
Divorced/separated		0.61	0.63	0.63	0.60	0.25
Widowed		0.15**	0.15**	0.15**	0.12***	0.14
Health status						
Good (ref)	-	-	(ns)	1.00*	(ns)	(ns)
Fair			1.67*	1.78*		
Poor			0.62	0.67		
LLSI	-	-	-	(ns)	(ns)	(ns)
Economic activity						
Retired/semi-retired (ref)	-	-	-	-	1.00***	1.00***
Employed					0.22***	0.20***
Self-employed					0.45	0.42
Permanently sick/disabled					0.46	0.86
Looking after home/family					2.61***	1.76
Relationship to person receiving care						
Spouse/partner (ref)	-	-	-	-	-	1.00***
Parent/parent-in-law						0.06***
Other relative/friend/neighbour						0.02***
Child/grandchild						0.15***
Constant	0.032***	0.026***	0.023***	0.021***	0.044***	10.418***
Log Likelihood ratio	1152.7	1109.0	1096.4	1095.6	1022.4	409.9

* p < 0.010

** p < 0.005

*** p < 0.001

(ns) = not significant.

[a] Note: 'Unemployed' was excluded from the models, as there were no unemployed among the round-the-clock carers.

Source: ELSA, Wave 3

Discussion

Some of the results presented in this article are in line with existing research on informal care provision in older ages, while other findings raise additional policy-relevant questions which require further investigation. The results suggest that women are more likely than men to provide care overall and more likely to care across all age groups, except among the oldest old. Women also appear more likely than men to be combining different caring roles and to be caring for more than one person; however, men are more likely than women to be caring for their spouses/partners.

The results of this article suggest the relationship between the provision of care and health status over the age of 50 is more difficult to assess. The greater likelihood among carers to report good health -compared to non-carers- may be explained by the fact that a certain level of good health is necessary in order to be able to care for another person. The prevalence of a good health status appears to be declining as care intensity increases; however, round-the-clock carers seem to report better health than some of the less intense carers. The area of health status in the latter part of the life course is as dynamic as the provision of care. For example, an older person may have been providing care to their grandchild in 2002, but by 2006 they may be receiving care from their children and providing support within the household to their spouse. Further research could investigate the relationship between care provision and health status, but also care receipt and health status, over time, using the ELSA dataset for longitudinal analysis.

It was found that round-the-clock carers constituted more than one fifth of all carers aged 50 and over, and that, compared to carers who provide care for less than 168 hours per week, round-the-clock carers are far less likely to be engaged in economic activity, but more likely than less intense carers to report a good health status. The logistic regression results further confirmed that care provision by people aged 50 and over is a profoundly gendered experience, as the determinants of being a round-the-clock carer differ for men and for women, except for the effect of the relationship to the person receiving care. Further research could investigate the factors that affect the chances of being a round-the-clock carer in greater depth and, more crucially, over time. Understanding the socio-economic position of carers better can contribute to additional policy-relevant mechanisms of recognising and supporting carers, particularly of those caring round-the-clock.

Key Findings

- Women were more likely to care than men, and ‘younger old’ people (50-64) more likely to care than ‘older old’ people (65 and over).
- Over half of all carers provided up to 19 hours of care per week, and just over one fifth provided round-the-clock care (168 hours). Men were more likely than women to care for up to 19 hours, and less likely to care between 20-49 hours.
- Men and women were equally likely to care round-the-clock, and the majority of these carers looked after their spouse/partner. However, women were more likely to be round-the-clock carers at a younger age and to be looking after parents/parents-in-law or children/grand-children, while men were more likely to be round-the-clock carers at older ages and to be

Key Findings

looking after their spouse/partner.

- Patterns of care provision varied across old age and in different ways for men and women. Women were more likely than men to care across their older ages, more likely to care for longer hours, and more likely to care for all categories of care recipients except for spouses/partners. Men were more likely to care in older ages and for their spouse/partner.
- More than three quarters of carers were married, and two fifths of all carers lived with the person receiving care.
- Carers appeared to be more likely than non-carers to report good health and not to report a limiting longstanding illness, but health status deteriorated as care intensity rose. The exception was round-the-clock carers, who reported better health status than some of the less intense carers.
- Men aged 50-64 were more likely to be economically active than women, both among carers and non-carers. Economically active persons were less likely to provide care compared to economically inactive persons, and the prevalence of economic activity decreased as care intensity rose. However, one quarter of round-the-clock carers in this age group were economically active, and the majority of these were employed.
- The strongest predictors of being a round-the-clock carer for both men and women are the relationship to the person they care for and their economic activity status. Among women, marital status also remains an important determinant, while age appears to be a more important determinant for men.

Acknowledgements

The author wishes to thank Maria Evandrou, Jane Falkingham and the anonymous referees for their useful comments on earlier drafts of this article. All responsibility for the analysis and interpretation of the data rests with the author.

References

1 Office for National Statistics (2003) *Census 2001: Key Statistics for local authorities in England and Wales.* London, The Stationery Office.

2 Maher, J and Green, H (2002) *Carers 2000*. London: The Stationery Office.

3 See for example: Evandrou, M and Glaser, K (2003) ‘Combining work and family life: the pension penalty of caring’. *Ageing and Society* 23 (5) 583-602.

4 Arber, S (2006) 'Gender and Later Life: Change, Choice and Constraints'. In: Vincent, JA, Phillipson, CR and Downs, M :*The Futures of Old Age*. London: Sage and the BSG. Pp 54-61.

5 National Population Projections, 2008-based (2009) Office for National Statistics, Statistical Bulletin. Available at: www.statistics.gov.uk/statbase/Product.asp?vlnk=8519

6 Pickard, L, Wittenberg, R, Comas-Herrera, A, Davies, B and Darton, R (2000) 'Relying on informal care in the 21st century? Informal care for elderly people in England to 2031'. *Ageing and Society* 20: 745-772.

7 See for example: Arber, S and Ginn, J (1990) 'The Meaning of Informal Care: Gender and the Contribution of Elderly People'. *Ageing and Society* 10: 429-454.

8 Dahlberg, L Demack, S and Bambra, C (2007) 'Age and gender of Informal Carers: A Population-based Study in the UK'. *Health and Social Care in the Community* 5: 439-445.

9 Office for National Statistics (2006) *Focus on Health*. 2006 Edition. Basingstoke: Palgrave Macmillan. Available at: www.statistics.gov.uk/statbase/Product.asp?vlnk=12985

10 See for example: Arber, S and Ginn, J (1995) 'Gender Differences in Informal Care'. *Health and Social Care in the Community* 3: 19-31.

11 Glaser, K and Grundy, E (2002) 'Class, caring and disability: evidence from the British Retirement Study'. *Ageing and Society* 22: 325-342.

12 Hyde, M and Janovic, M (2004). 'Social activity'. In: Marmot, M *et al.* (eds) *Health, wealth and lifestyles of the older population in England. The 2002 English Longitudinal Study of Ageing*. London: The Institute for Fiscal Studies. Pages 167-179.

13 Evandrou, M and Glaser, K (2004) 'Family, work and quality of life: Changing economic and social roles through the lifecourse'. *Ageing and Society* 24 (1) 1-21.

14 Buckner, LJ and Yeandle, SM (2007) *Valuing Carers: Calculating the Value of Unpaid Care*, Research Report Prepared for Carers UK. London: Carers UK.

15 ELSA Technical Reports (Waves 1,2,3 and 4 – Phase 1). Available from the Economic and Social Data Service: www.esds.ac.uk

16 Del Bono, E, Sala, E and Hancock, R (2009) 'Older carers in the UK: are there really gender differences? New analysis of the Individual Sample of Anonymised Records from the 2001 UK Census'. *Health and Social Care in the Community* 17 (3) 267-273.

17 Young, H, Grundy, E and Jitlal, M. (2006) *Care Providers, Care Recievers. A Longitudinal Perspective.* York: Joseph Rowntree Foundation.

18 More than 70 per cent of all married persons in Wave 3 reported good health, compared to 60 per cent of non-married persons, and conversely, six per cent of all married persons reported poor health, compared to 10 per cent of non-married persons ($p<0.001$).

Migration in later life: evidence from the British Household Panel Study

Maria Evandrou, Jane Falkingham, Marcus Green

ESRC Centre for Population Change & Centre for Research on Ageing, University of Southampton

Abstract

This article uses data from the British Household Panel Study over the period 1991 – 2007 to examine the factors associated with residential mobility among people aged 50 and over. In line with earlier research, the likelihood of migrating, that is, changing address, is found to vary according to the demographic and socio-economic characteristics of the older person. Those in late middle age (50-59) and the oldest-old (90 and over) were most likely to move. Migration was also strongly associated with changes in partnership, health and economic status during the last 12 months, highlighting the importance of seeing migration within a life course context with certain life course events such as divorce, widowhood or retirement being important triggers for prompting a move. As divorce and remarriage become more common in later life, 'relationship driven migration' is likely to become more important, adding a new category to the classical typology of later life migration.

Contents

Introduction

The UK's population is ageing. According to the ONS 2009 mid-year population estimates published in June 2010, 21.3 million people are currently aged 50 and over[1]. A better understanding of the characteristics of this section of the population is critical for the efficient and effective planning of services both by the public and private sector. Most migration studies focus on labour migrants and there is relatively little recent research on older migrants. Much of the existing research on migration in later life has focused on international migration *from* England and Wales to foreign destinations in retirement, most notably to the countries of southern Europe (Williams *et al* [2]; King *et al* [3]; Warnes and Patterson[4]; King *et al* [5]); while more recently Green, Evandrou and Falkingham (2009)[6] have examined international migration to the UK in later life. However, the vast majority of moves in later life are moves within the UK. The recent study of migration trends at older ages by Uren and Goldring[7] found that around a third of people aged 50 and over in England and Wales had changed address during the period 1991-2001.

Litwak and Longino (1987)[8] identified three distinct types of migration at older ages: 'amenity' related or retirement migration undertaken to improve quality of life; disability or health-related migration, undertaken to move closer to kin; and finally moves into institutions towards the end of life. They highlight that the balance between these types of migration will vary across the life course, with migratory moves being intimately related to other life course events. Thus the characteristics of those undertaking life-style moves at retirement may be expected to vary significantly from those moving in later old age.

This article aims to explore the characteristics of older migrants (aged 50 and over) in the UK between 1991 and 2007 using data from the British Household Panel Survey (BHPS). It builds upon previous work by Grundy[9 10], Glaser and Grundy[11] and Uren and Goldring[7] both by extending the analysis to include a wider range of socio-economic characteristics and by investigating annual moves rather than moves in a ten year period. The availability of a range of data on *annual* transitions means that migratory moves may be more explicitly linked to the timing of other life course events such as divorce, widowhood or retirement. Older migrants are defined as those people who have changed residential address during the reference period. As such the article looks at all residential mobility rather than the more narrowly defined internal migration across administrative boundaries. The article is limited to those older people living in the community and does not explore moves into institutional care.

Data Sources

Data on internal migration, or residential mobility, in the UK are often limited as there is no legal obligation to register a change of address. Data from patient registers can be used to provide some estimates of the extent, or quantum, of internal migration[12], but is restricted in the amount of detail available on the characteristics of the migrants themselves. Previous research on older migrants has tended to use data either from the census[6 13], or from the ONS Longitudinal Study[7 11]. Although the latter has the advantage of large sample sizes, being based on a one per cent sample of the census in England and Wales, data on the characteristics of the respondents is only available at ten yearly intervals and is limited to the information collected in the census, which excludes income and other indicators of financial well-being. In recent years national surveys such as the Labour Force Survey and General Household Survey, now the General LiFestyle Survey (GLF), which

collect detailed information on a wider range of socio-economic characteristics have included questions on change of address in the last year. However, until recently, they have only provided cross-sectional data and do not allow investigation of migration and its relationship with other life course events. Moreover, in any given year, only a small number of older people sampled will have migrated, limiting the usefulness of these sources for examining migration in later life.

This article attempts to overcome these limitations by using pooled data from the British Household Panel Study (BHPS) over the period 1991 – 2007. The BHPS is a longitudinal, multi-purpose study carried out by the UK Longitudinal Studies Centre and offers a wide array of variables to explore the profile of older migrants. As the data is longitudinal, it is possible to investigate how life course events, such as experiencing a divorce or being widowed, are associated with a migratory move in the same period. It is important to note that migration here refers to any change of address. This differs from standard approaches; demographers often think of migration as only moves between statistical areas, with within-area movement being termed 'residential mobility'. In this article, the terms 'migration', 'residential mobility' and 'change of address' are used interchangeably.

In order to overcome the problem of migration being a relatively rare event at older ages, observations from all 17 waves of the BHPS are used to create a pooled dataset. Respondents in adjacent waves are matched using the cross-wave person identifier to create a series of subsets containing the characteristics of the respondents at two points in time (t1 and t2). For example, respondents in 1991 are matched with their responses in 1992 to create a subset for 1991-92, data from 1992 are matched with 1993, and so on, until 2006-07. The resulting 16 paired year subsets are then merged to create the final dataset. Respondents who only provide one interview at one of the two successive waves are necessarily excluded. For the purposes of this analysis, the sample is then restricted to respondents aged 50 and over at t1. This yielded a total sample size of 71,356 people aged 50 and over, of whom 2,397 (3.4 per cent) have migrated between t1 and t2. Of these 335 are serial migrants, as they have changed address both between t0 and t1, and then again between t1 and t2.

The BHPS is a panel survey which follows the same individuals over time, this means that the same individuals may appear in the pooled paired year dataset more than once. For example, an individual who was aged 75 in 1991, and who was interviewed in each of the first six waves of the panel, will appear in the dataset five times. This means that the results may be confounded by correlation within cases over time[14]. However, this disadvantage is offset by the greater statistical power and reliability of estimation afforded by the larger sample size.

Identifying migrants

The BHPS collects information on respondent's address one year ago. It is therefore possible to identify migrants in several ways. In wave 1, those older people who answered that they were not resident at their current address one year prior to the interview, may be thought of as having migrated between t0 and t1. Similarly in wave 2, those who answered that they were not resident at their current address one year prior to the interview, may be thought of as having migrated between t1 and t2. Finally, those who answered that they were not resident at their current address one year prior to the interview in both wave 1 and 2, may be thought of as experiencing serial migration.

As this article is primarily interested in understanding the characteristics that are associated with a move in order to be better able to forecast local population needs, older migrants are defined here

as those who move between t1 and t2. This allows us to investigate which demographic and socio-economic characteristics at t1 are associated with a move in the *following* 12 months (interviews between waves are usually 12 months apart). As the data are limited to people who are resident in households at the time of the BHPS survey, migration into institutions is not observed. The inclusion of the variable 'survey year' in the dataset allows investigation of period effects, such as changes in policy, economic climate and political context.

Life course events

Previous research has shown that migration behaviour is closely associated with other life course events[8]. The rich longitudinal data available from the BHPS allows us to construct a series of variables that capture key life course transitions, such as getting married, divorced or widowed, becoming unemployed or retiring or experiencing deteriorating (or improving) physical and mental health.

A series of variables capturing these key life transitions are derived as follows:

- Newly partnered (not partnered at t1 and partnered (defined as married, cohabiting or in civil partnership) at t2)
- Newly divorced/separated (partnered at t1 and divorced/separated at t2)
- Newly widowed (partnered/separated at t1 and widowed t2)
- Newly unemployed (employed at t1 and unemployed at t2)
- Newly self-employed (employed or unemployed at t1 and self-employed at t2)
- Newly retired (economically active at t1and retired at t2)

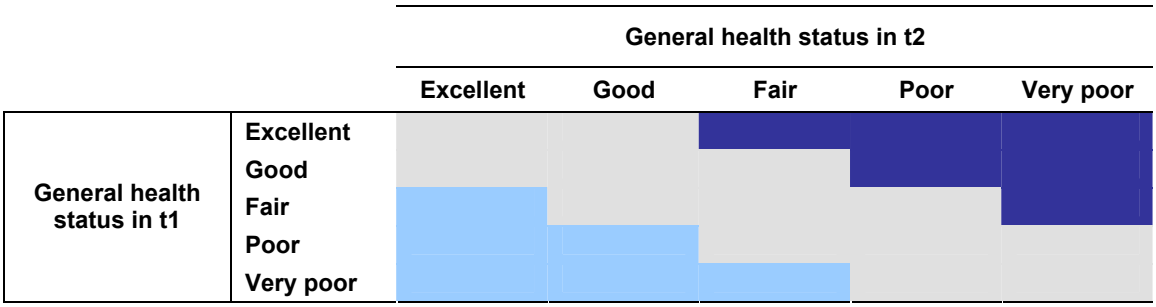
It is important to note that we do not have information on the exact timing of these life course events and how they relate to the timing of migration during the year. If migrations are spread evenly across the year, some will have occurred only weeks or even days before the interview at t2. Similarly, the changes in marital status or economic activity may have occurred at any stage between the two interviews. The variables do not allow us to investigate causation, but simply to capture life course events that occur in the same period as a change in address.

Health

The BHPS contains a number of measures of health including general health status over the past year (measured using a five response Likert scale of excellent, good, fair, poor or very poor) and whether a person suffers from a long-term illness that limits their activity.

Given that data are available at two points in time, it is possible to create a variable that captures changes in self-reported health status across time. Here an improvement in health is defined as experiencing an upward movement of *more than one* category in self-reported health status between t1 and t2; that is, a move from fair to excellent, or a move from poor to good or excellent, or a move from very poor to fair, good or excellent (shaded pale blue in **Figure 1**). Similarly, a deterioration in health is defined as experiencing a downward movement of more than one category between t1 and t2; that is from excellent to fair, poor or very poor, or from good to poor or very poor, or from fair to very poor (shaded dark blue). Health is defined as staying the same if the respondent reports the same health status, plus or minus one category, at both points in time (shaded grey).

Figure 1 Five response Likert scale for measuring change in health status over time with three outcomes



Results

Table 1 shows the percentage of older people who changed address between two successful waves of the BHPS over the period 1991-2007. Overall 3.4 per cent of people aged 50 and over migrated between two waves and there is no overall difference by gender. However, there are clear gender differences by age, with the migration level being significantly higher for women than for men at ages 80-84 and 85-89 (see **Figure 2**). This may in part reflect sex differentials in mortality and the higher likelihood of women being widowed and then moving at these ages.

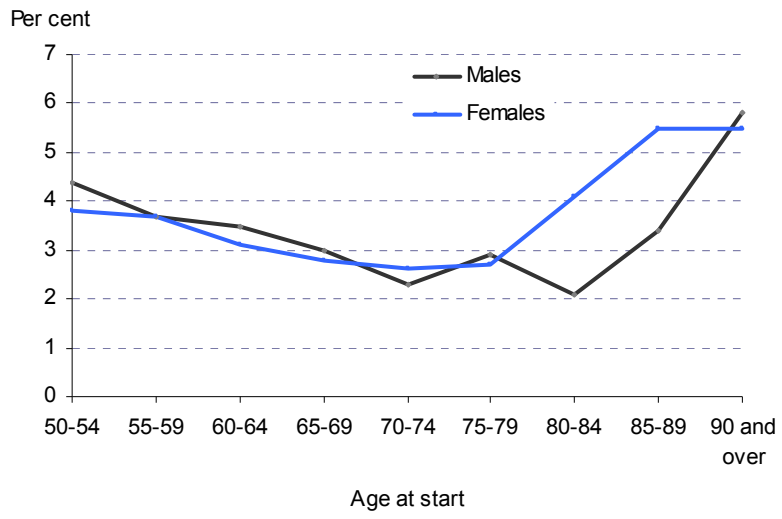
Table 1 People aged 50 and over who migrated between waves, by age at start of the period, UK

Age group	Males		Females		All	
	Per cent migrating	Number in sample	Per cent migrating	Number in sample	Per cent migrating	Number in sample
50-54	4.4	6,618	3.8	7,809	4.1	14,427
55-59	3.7	5,927	3.7	6,978	3.7	12,905
60-64	3.5	5,030	3.1	5,937	3.3	10,967
65-69	3.0	4,595	2.8	5,413	2.9	10,008
70-74	2.3	4,127	2.6	5,187	2.4	9,314
75-79	2.9	2,862	2.7	4,137	2.8	6,999
80-84	2.1	1,662	4.1	2,725	3.3	4,387
85-89	3.4	644	5.5	1,190	4.8	1,834
90+	5.8	154	5.5	361	5.6	515
All ages	**3.4**	**31,619**	**3.4**	**39,737**	**3.4**	**71,356**

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

The relationship between migration and age is U-shaped. For both men and women the likelihood of migrating decreases with age until the late 70s or early 80s, at which point the likelihood starts to increase. This is a similar finding to that of Conway and Rork in their study of old age migration in the USA[15]. They also argued that that those moving at younger ages tend to do so on the grounds of amenity, that is, moving to improve life style, whereas those who move at oldest-old ages move because they may need assistance in terms of informal or formal care provision. As the BHPS data excludes moves into institutions, the percentage migrating at ages 80 and over is likely to be considerably higher than the percentages shown in Figure 2.

Figure 2 People aged 50 and over migrating between waves by sex and age group, percentage, UK



Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Partnership status

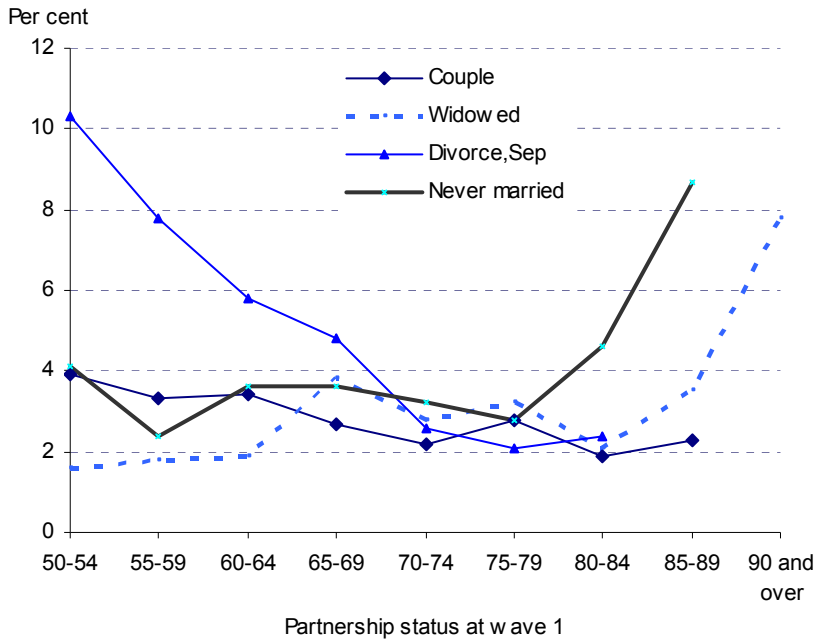
Previous analyses have shown that changes in partnership status may influence an individual's decision to migrate in later life[7 8 11]. **Figure 3a** and **3b** highlight the differences in the likelihood of experiencing a migratory move according to partnership status at the start of the period by age. Living as a couple includes those who are legally married, or in a civil partnership, as well as cohabiting couples. At ages under 70 the proportion migrating in the coming year, that is between t1 and t2, is much higher among those men, who report being divorced at the start of the period, compared to all other groups, while at older ages, it is those men who are widowed, who are most likely to experience a move. This pattern is also evident among women, although the differentials by partnership status are less marked. **Table 2** sheds further light on the relationship between partnership status and migration by separating those older people who experience a change in their relationship status between successive interviews from those who do not.

Table 2 People aged 50 and over who migrated between waves, by partnership status at the end of the period, UK

	Males		Females		All	
	Per cent migrating	Number in sample	Per cent migrating	Number in sample	Per cent migrating	Number in sample
Newly partnered	25.0	164	17.5	177	21.1	341
Continuing couple	2.9	24,323	2.8	22,850	2.9	47,173
Newly widowed	5.6	324	5.2	688	5.3	1,012
Continuing widowed	3.0	2,777	3.7	10,176	3.6	12,953
Newly divorced/separated	21.6	171	11.3	301	15.0	472
Continuing divorced/separated	5.6	1,781	4.8	3,272	5.1	5,053
Never married	3.4	2,068	2.9	2,263	3.1	4,331
All	**3.4**	**31,608**	**3.4**	**39,727**	**3.4**	**71,335**

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

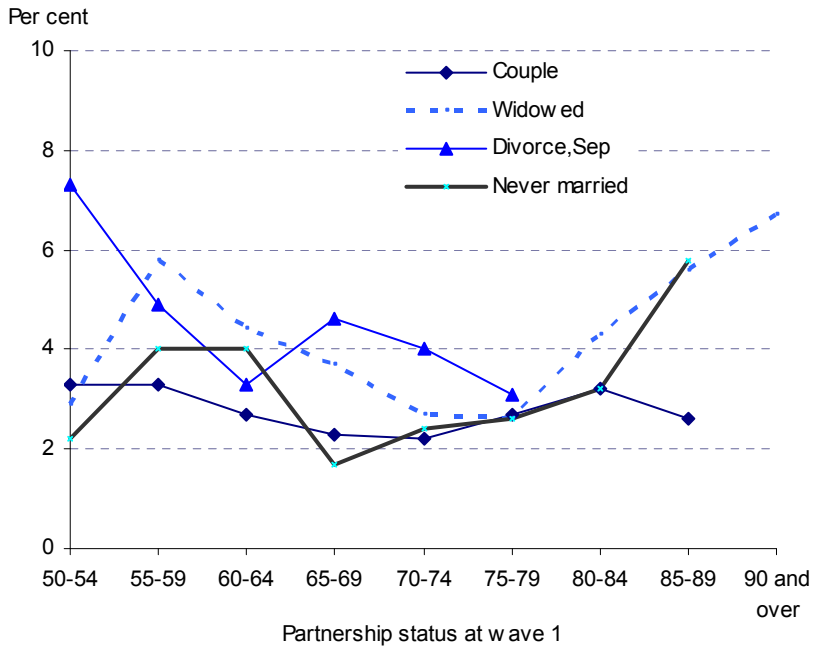
Table 3a Men migrating between waves by age and marital status



Note: Data points where N < 20 are excluded

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Table 3b Women migrating between waves by age and marital status



Note: Data points where N < 20 are excluded

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Migration is highest among those older men and women who have formed a new partnership, with 25 percent of men and 17.5 percent of women who have married, started cohabiting, or entered

into a civil partnership between t1 and t2, also changed address during the year. Experiencing a breakdown in relationship is also strongly associated with a move, with 21.6 percent of men and 11.3 percent of women, who divorced or separated from a partnership between waves, also changed their place of residence[16]. There is also clear evidence of a relationship between recent widowhood and migration, with those who were newly widowed being almost twice as likely to move as those who were widowed previously. This is similar to results obtained in the USA using data from the Panel Study of Income Dynamics[17].

Table 2 also provides some evidence that the impact of relationship breakdown on the likelihood of migrating persists beyond the first year, with those older men and women who are 'continuing divorcees or separatees' being nearly as likely to migrate as those who are newly widowed. Interestingly, six per cent of those who are newly partnered and two per cent of those who are newly divorced/separated are serial movers (that is, they changed address both between t1 and t2 and also in the previous year between t0 and t1). This compares with just 0.5 percent of all people aged 50 and over, which further reinforces the notion that changes in relationship are an important trigger for residential mobility. As divorce and remarriage become more common in later life, it is perhaps time for the classic Litwak and Longino[8] typology of old age migration to be extended by the inclusion of a fourth category, that is 'relationship driven migration'.

Socio-economic characteristics

The relationship between socio-economic circumstances and migration in later life is not clear cut. Studies in North America by Meyer and Speare[18] and Moore and Rosenberg[19] found that individuals with fewer financial constraints were more likely to migrate than those older people on lower incomes. In contrast, Silverstein[20] found that those with lower incomes were more likely to reduce the temporal distance between themselves and their children. These types of moves have been branded by Conway and Rork[15] as 'assistance moves', where individuals move towards family or other support systems. Clark and White[21], focusing on the relationship between income and mobility among older people living in the rented sector, found that older renters were more likely to migrate if they had low or high income, as opposed to middle income. Similarly Walters[22] and Wiseman[23] have found that older migrants were more likely to have, either above average incomes (amenity migrants), or lower incomes (assistance migrants).

Analysis of the BHPS data found no significant difference in the level of migration for those aged over 50 according to the relative income position of the older person (annual income was recoded into quintiles *within* years to avoid problems of inflation over time). The only exception to this was for those aged 60-69, where people who were ranked in the bottom quintile, that is the poorest 20 percent, were less likely to move (2.4 per cent) than the age group as a whole (3.1 per cent). This suggests that financial position may play a role in facilitating moves around the time of retirement, but does not play such an important role in the UK at other stages. However, that is not to say that economic factors are unimportant.

Table 3 shows migration percentages for men and women aged 50 and over by their economic activity status at t1. Those men and women, who are unemployed at the start of the period, are significantly more likely to migrate during the following 12 months than those who are employed. So too are those who are self-employed or long term sick or disabled, while those who are retired are less likely to migrate.

Table 3 People aged 50 and over who migrated between waves, by economic activity status at the *start* of the period, UK

Economic activity status	Males		Females	
	Per cent migrating	Number in sample	Per cent migrating	Number in sample
Employed	3.4	9,513	3.2	10,198
Self-employed	4.3	3,386	5.0	1,062
Unemployed	4.9	836	6.6	471
Retired	2.9	15,357	3.2	21,097
Long term sick/disabled	4.5	2,157	4.9	1,920
Other (incl. keeping house, FT student)	3.8	370	3.0	4,989
All	**3.4**	**31,619**	**3.4**	**39,737**

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Again it is interesting to investigate whether *changes* in economic activity status *over the year* are associated with a move during that year. **Table 4** therefore distinguishes between those who continue in the same economic activity status and those who have newly acquired that status. In all cases, with the exception of women who are unemployed or long term sick, migration among 'new entrants' is higher than among 'stayers'. This again demonstrates the inter-dependency of different life course events, with residential moves often being associated with changes in employment status.

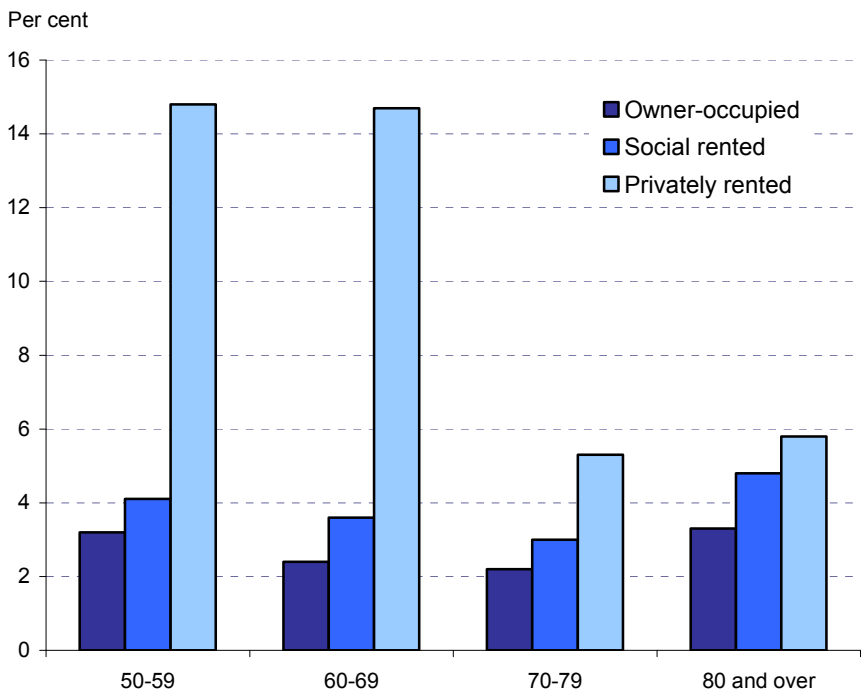
Table 4 People aged 50 and over who migrated between waves by economic activity status at the *end* of the period, UK

Economic activity status	Males		Females	
	Per cent migrating	Number in sample	Per cent migrating	Number in sample
Newly employed	4.0	683	4.1	652
Continuing employed	3.1	8,252	2.8	8,857
Newly self employed	6.8	395	4.6	239
Continuing self employed	3.7	2,847	3.1	775
Newly unemployed	6.6	364	6.7	283
Continuing unemployed	3.8	365	9.9	121
Newly retired	5.3	1,543	5.0	2,759
Continuing retired	2.8	14,736	3.2	19,512
Newly long term sick	5.8	591	4.6	627
Continuing long term sick	4.3	1,598	4.7	1348
Other	6.3	254	3.1	4,564
All	**3.4**	**31,619**	**3.4**	**39,737**

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Housing tenure is often used as an indicator of socio-economic status in the UK. Previous studies have found tenure to be strongly related to the propensity to migrate in later life[7 9 10]. **Table 5** illustrates that among people aged 50 and over, migration is highest for those living in privately rented accommodation, and especially for the minority renting in the furnished private sector (19 per cent). **Figure 4** confirms for all age groups that those living in privately rented accommodation are the most likely to move, although the differential according to tenure decreases with age.

Figure 4 People aged 50 and over migrating by household tenure and age group, percentage, UK



Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Table 5 People aged 50 and over who migrated between waves by housing tenure at the start of the period, percentages, UK

Tenure type	Males	
	Per cent migrating	Number in analysis
Owners outright	2.5	36,323
Owns with a mortgage	3.3	16,801
Local authority rented	3.3	11,483
Housing association rented	5.0	2,799
Rented from Employer	11.1	488
Other privately rented unfurnished	9.7	2,248
Other privately rented furnished	19.2	646
All	**3.4**	**31,619**

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Health

In a similar fashion to the relationship between migration and wealth discussed above, the relationship between health and migration has also been postulated as being U-shaped[24]. It is argued that good health may help facilitate amenity driven moves in early old age, while poor health may act as a trigger for an assistance move at older ages.

Figure 5a and **5b** confirm the positive association between migration and poor health at older ages, with much higher migration for men and women with a limiting long term illness at ages 80 and over. However, migration levels for those with a limiting long term illness *are* also higher than for those who do not suffer from a long term illness at younger ages. This suggests that it is too simplistic to assume that all migration at younger old ages (50-69 years) is amenity driven, but rather, that migration for these age groups may be a mixture of both amenity and health related migration. This picture is further supported by **Table 6**, which shows migration percentages by self reported health status by age group. There is little evidence of a U-shaped relationship, but rather, that the percentage migrating tends to rise for all age groups as health worsens.

Table 6 Percentage of men and women migrating by self reported health status at the start of the period and age group, UK

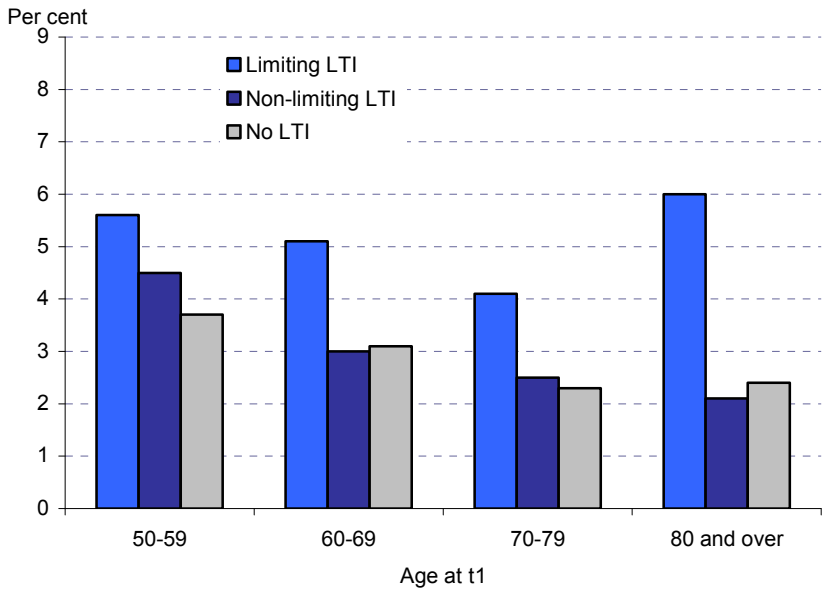
Health status over the last 12 months at t1	Aged 50-59	Aged 60-69	Aged 70-79	Aged 80-89
Men				
Excellent	4.5	3.3	2.0	2.2
Good	3.7	2.7	2.1	2.1
Fair	3.5	3.6	2.6	3.9
Poor	4.5	3.8	3.9	3.6
Very poor	9.5	6.3	5.0	-
Women				
Excellent	4.0	2.7	2.1	4.0
Good	3.1	2.8	2.2	4.2
Fair	4.3	2.9	3.2	4.6
Poor	5.0	3.7	3.7	6.3
Very poor	3.4	3.5	2.8	7.3

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

As with partnership status and economic activity, the BHPS data allows us to look at how *changes* in health are related to migration. As discussed above, an improvement in health was defined as experiencing an upward movement of *more than one* category in self reported health status between t1 and t2, while a deterioration in health was defined as experiencing a downward movement of more than one category between t1 and t2. Health was defined as staying the same if the respondent reported the same health or plus or minus one category.

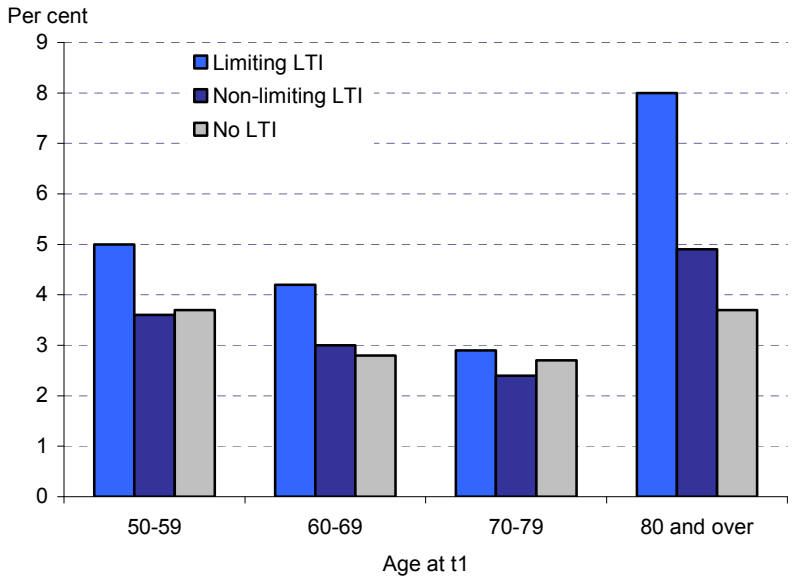
In contrast to previous studies[24], **Figure 6** shows that at ages 50-69 men and women whose health has deteriorated between waves have similar migration levels to those whose health has improved, while at ages 70-79 migration is highest among those whose health has improved. The direction of causation is not clear. On the one hand changing one's place of residence may result in improved subjective well-being and therefore self-perceived health. Alternatively, improved health may have facilitated the migratory move. Whatever is the case, this highlights that the relationship between health status and migration is more complex than at first sight.

Figure 5a Men aged 50 and over migrating by having a limiting long term illness and age at t1



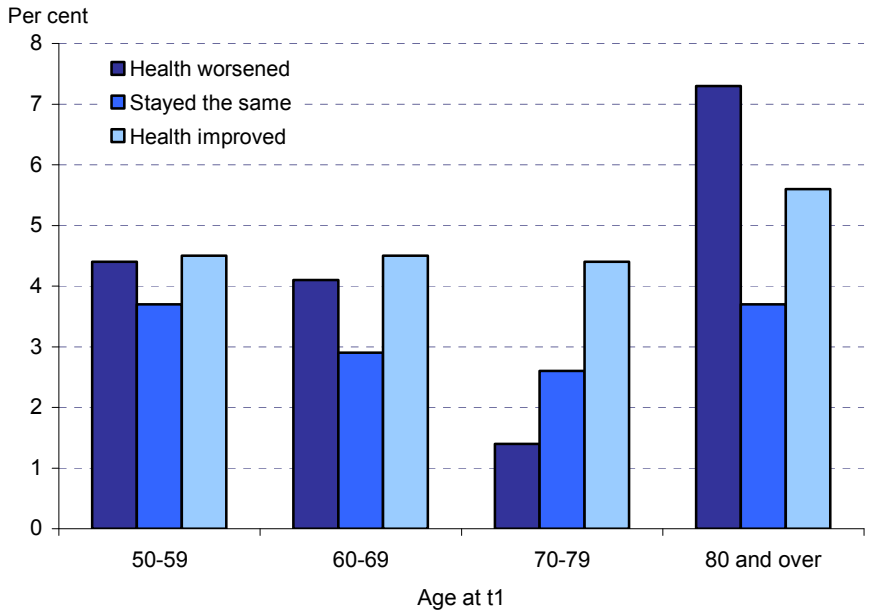
Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Figure 5b Women aged 50 and over migrating by having a limiting long term illness and age at t1



Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Figure 6 **People aged 50 and over migrating by change in health status (t1- t2) and age group at t1**



Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

Migration and life course transitions

So far, we have examined the relationship between migration and older individuals' socio-economic characteristics and life course transitions separately. **Table 7** explores the extent to which these factors remain significant after controlling for other things. Living in rented accommodation remains significantly associated with increased mobility in later life, indicating that those older people in the private rented sector are more able to change their housing options than owner occupiers.

Changes in partnership status also remain highly significant, with those older people who enter into a new partnership being over six times more likely to move during the year compared with those who are never married. Conversely, those who divorce or separate are nearly five times as likely to move. Previous research has highlighted widowhood as a key risk factor and those who are newly widowed are twice as likely to move as the never married. However, there is no significant elevated risk of migrating among the continuing widowed. The heightened likelihood of moving among older people who are divorced, does however, persist.

Retirement is also an important risk factor for moving, but so too is exit from the labour market through unemployment and/or becoming self employed. Fielding's 'escalator region' work in the early 1990s using ONS Longitudinal Study 1971-1981 highlighted the distinctive migration of older people into self-employment (or the 'petite bourgeoisie' in his vocabulary), especially in moves from South East England to adjacent regions[25], but these factors have received little attention in the more recent literature.

Table 7 Odds ratios of migrating between t1 and t2, persons aged 50 and over, 1991-2007, UK

	Odds ratio	95% Confidence interval
Age at t1* **		
50-54	1.00	
55-59	0.87 *	0.76-0.99
60-64	0.69***	0.59-0.81
65-69	0.58***	0.48-0.71
70-74	0.46***	0.37-0.57
75-79	0.50***	0.40-0.62
80-84	0.56***	0.44-0.71
85-89	0.73 *	0.54-0.97
90+	0.93	0.60-1.46
**General health status over last 12 months at t1* **		
Excellent	1.00	
Good	0.89	0.78-1.01
Fair	1.00	0.87-1.14
Poor	1.07	0.90-1.28
Very poor	1.16	0.89-1.50
LLTI at t1* **		
No limiting long term illness	1.00	
Limiting LTI	1.41***	1.23-1.62
Non-limiting LTI	1.06	0.96-1.18
Partnership status at t2* **		
Never married	1.00	
Newly partnered	7.64***	5.45-10.69
Continuing couple	1.01	0.83-1.22
Newly widowed	1.89***	1.33-2.68
Continuing widowed	1.28 *	1.03-1.59
Newly divorced, separated	5.41***	3.90-7.52
Continuing divorced, separated	1.52 **	1.21-1.91
Economic status at t2* **		
Other	1.00	
Newly employed	1.24	0.88-1.73
Continuing employed	0.84	0.69-1.03
Newly self employed	1.64 *	1.10-2.46
Continuing self employed	1.00	0.77-1.30
Newly unemployed	1.70 **	1.17-2.47
Continuing unemployed	1.24	0.79-1.94
Newly retired	1.73***	1.38-2.17
Continuing retired	1.20	0.98-1.46
Newly long term sick	1.31	0.95-1.82
Continuing long term sick	0.93	0.71-1.22
Tenure at t1* **		
Owned	1.00	
Social rented	1.19 **	1.06-1.33
Privately rented	4.05***	3.55-4.62

Variables entered in forward conditional step wise model in the following order: tenure, partnership status, economic status, age, limiting long term illness, general health status.

Change in health status, sex and year were not significant

*** p<0.001 ** p<0.01 * p<0.05

Source: authors' own analysis of pooled paired year BHPS data 1991-2007.

The U-shaped relationship with age also remains, with the likelihood of moving being highest at ages 50-54 once health and widowhood (two factors strongly associated with moving at older ages) are taken into account. Interestingly, the two health variables were the last of the variables to be selected into the model, despite the emphasis in the literature on health as a determinant of migration in later life. In part this reflects the fact that this study is restricted to moves within the community and thus excludes moves into residential care. The BHPS does contain information on those who are lost to follow up through moves into institutions where the household remains in the survey, and future work will attempt to explore this further.

Conclusion

Typologies of migration patterns in later life have largely concentrated on classifying moves into two broad groups, those associated with improved lifestyle around the time of retirement and moves later in life associated with failing health. Our research has confirmed that both are important factors associated with moving at ages post 50. However, focussing on these two types of moves risks missing a key part of the picture. In particular, migration in later life – as is also the case at younger ages - is strongly associated with the formation and dissolution of partnerships. Moreover, the impact of partnership dissolution has a persisting effect with older people who are divorced or separated being more likely to move than the never married even after the first year of partnership breakdown. As re-partnering becomes increasingly common in older ages, it will be important that policy makers and planners take this into account.

Key findings

- Around 3.5 per cent of all people aged 50 and over living in private households move addresses each year and this figure has been relatively constant across time.
- The likelihood of migrating decreases with age until the late 70s/early 80s when the likelihood rises.
- Those older people living in rented housing are more likely to move than those living in owner occupied housing, with those living in the private rented sector experiencing higher migration than those living in the social rented sector.
- The likelihood of migrating in later life is strongly linked to other life course events:
  - Forming a new partnership or experiencing a breakdown in an existing partnership, either through divorce or widowhood, can be an important trigger for a move.
  - Exit from the labour market through retirement or unemployment is also associated with migration.
- Existing typologies of migration in later life may need to be adapted to reflect changes in social norms as re-partnering at older ages becomes more common.
- Poor health may also trigger a move. However, the relationship between migration and health is not straightforward with higher migration at ages 70-79 among those whose health has improved during the year in which the move has taken place.

Acknowledgements

Maria Evandrou and Jane Falkingham are co-Directors of the ESRC Centre for Population Change (CPC) funded by ESRC Grant number RES-625-28-001. Marcus Green is a PhD student within the Centre. The Centre for Population Change (CPC) is a joint initiative between the University of Southampton and a consortium of Scottish Universities in partnership with ONS and GROS. Maria Evandrou is the Director of the Centre for research on ageing at the University of Southampton.

The authors are grateful for the constructive comments from the anonymous reviewers. The findings, interpretations, and conclusions expressed in this article are entirely those of the authors and should not be attributed in any manner to ONS or GROS. Data from the British Household Panel Survey (BHPS) were supplied by the UK Data Archive. Neither the original collectors of the data nor the archive bear any responsibility for the analysis or interpretations presented here.

References

1 Office for National Statistics (2010) 2009 Mid-year population estimates. Available at: www.statistics.gov.uk/statbase/Product.asp?vlnk=15106

2 Williams, AM, King, R and Warnes, AM (1997) 'A Place in the Sun; International Retirement Migration from Northern to Southern Europe'. *European Urban and Regional Studies,* Vol 4 (2) 115-134.

3 King, R, Warnes, AM and Williams, AM (1998) 'International Retirement Migration in Europe'. *International Journal of Population Geography* 4: 91-111.

4 Warnes, AM and Patterson, G (1998) 'British Retirees in Malta: Components of the Cross-National Relationship'. *International Journal of Population Geography* 4: 113-133.

5 King, R, Warnes, AM and Williams, AM (2000) *Sunset Lives: British Retirement Migration to the Mediterranean*. Oxford; Berg.

6 Green, M, Evandrou, M and Falkingham, J (2009) 'Older International Migrants: who migrates to England and Wales in later life?' *Population Trends* 137: 33-40.

7 Uren, Z and Goldring, S (2007) 'Migration Trends at Older Ages in England and Wales'. *Population Trends* 130: 31-40. Available at: www.statistics.gov.uk/statbase/Product.asp?vlnk=6303

8 Litwak, E and Longino, C (1987) 'Migration Patterns Among the Elderly'. *The Gerontologist,* 7 (3) 266-272.

9 Grundy, E (1987) 'Retirement Migration and Its Consequences in England and Wales'. *Ageing and Society,* 7: 57-82.

10 Grundy, E (1993) 'Moves into Supported Private Households among Elderly People in England and Wales'. *Environment and Planning A,* 25: 1467-1479.

11 Glaser, K and Grundy, E (1998) 'Migration and Household Change in the Population Aged 65 and over, 1971-1991'. *International Journal of population Geography* 4: 323-339.

12 Scott, A and Kilbey, T (1999) ‘Can patient registers give an improved measure of internal migration in England and Wales? *Population Trends* 96: 44-45.

13 Warnes, AM (1996) ‘Migrations among older people’. *Reviews in Clinical Gerontology* 6:101-114.

14 Menard, S (2002) *Longitudinal Research*. Sage University Paper Series on Quantitative Applications in the Social Science no. 76. Thousand Oaks, CA, Sage.

15 Conway, K and Rork, J (2008) *Elderly and Non-Elderly Interstate Migrants – The Changing Roles of Socioeconomic, Disability and Veteran Status*. National Institutes of Health.

16 It is somewhat surprising that these figures are not higher, suggesting that a high proportion of older people who experience a partnership breakdown are still co-resident. There are several possible explanations. Firstly, one partner may have already migrated before the reference year when the divorce/separation was reported. Secondly, the BHPS may be biased towards those who remained in the family home, thus under-reporting on those who have moved as a result of breakdown. Further work is needed in order to identify both partners, in order to assess how many older men and women could not be traced following a breakdown.

17 Chevan, A (2005) ‘Holding On and Letting Go: Residential Mobility during Widowhood’. *Research on Aging,* 17 (3) 278-302.

18 Meyer, J and Speare, A (1985) ‘Distinctively Elderly Mobility: Types and Determinants’. *Economic Geography,* 61: 79-88.

19 Moore, E and Rosenberg, M (1994) ‘Residential Mobility and Migration among Canada’s Elderly’. *Aging: Canadian Perspectives*.

20 Silverstein, M (1995) ‘Stability and Change in Temporal Distance between the Elderly and their Children’. *Demography*, 32 (1) 29-45.

21 Clark, W and White, K (1990) ‘Modelling Elderly Mobility’. *Environment and Planning A,* Vol 22 (7) 909-924.

22 Walters, WH (2002) ‘Later-life migration in the United States: a review of recent research’, *Journal of Planning Literature* 17 (1) 37-66.

23 Wiseman, RF (1980)’ Why older people move: Theoretical issues’. *Research on Aging Vol 2*, 141-154.

24 Patrick, C (1980) ‘Health and Migration of the Elderly’. *Research on Aging Vol 2*, 233-241.

25 Fielding, AJ (1992) 'Migration and Social Mobility: South East England as an Escalator Region', *Regional Studies,* 26 (1) 1-15.

A comparison of official population projections with Bayesian time series forecasts for England and Wales

Guy J. Abel, Jakub Bijak and James Raymer

University of Southampton

Abstract

We compare official population projections with Bayesian time series forecasts for England and Wales. The Bayesian approach allows the integration of uncertainty in the data, models and model parameters in a coherent and consistent manner. Bayesian methodology for time-series forecasting is introduced, including autoregressive (AR) and stochastic volatility (SV) models. These models are then fitted to a historical time series of data from 1841 to 2007 and used to predict future population totals to 2033. These results are compared to the most recent projections produced by the Office for National Statistics. Sensitivity analyses are then performed to test the effect of changes in the prior uncertainty for a single parameter. Finally, in-sample forecasts are compared with actual population and previous official projections. The article ends with some conclusions and recommendations for future work.

Contents

Introduction

In recent years there has been an increasing emphasis by national statistical offices to include uncertainty in their official population projections so that the user community has a more realistic sense of what the future might hold. For most national statistical offices this has involved the inclusion of several plausible (deterministic) projection variants based on assumptions regarding future fertility, mortality and migration in a cohort-component population projection framework. In this article, we focus on the issues and practicalities of including uncertainty from a probabilistic viewpoint.

In the 1990s there were several convincing papers arguing for the need to move away from variant-style projections to probabilistic ones. See, for example, Ahlburg and Land 1992[1]; Lee and Tuljapurkar 1994[2]; Lutz 1996[3]; Bongaarts and Bulatao 2000[4]. The advantages are clear: probabilistic projections specify the likelihood that a particular future population value will occur given a set of assumptions about the underlying probability distributions. With variant projections, on the other hand, the user has no idea how likely they are. Here, the users have to trust that the experts have provided them with plausible scenarios representing the “most likely” (the principal projection) and the extremes (the high and low population projections). Of course, in both cases, the quality of the forecasts depends on the input data, projection model and assumptions made.

Despite the advantages of a probabilistic approach and the abundance of applications[5], nearly all national statistical offices in the world still rely on deterministic variant projections to provide uncertainty[6]. However, progress is being made; the Office for National Statistics (ONS), for example, has recently been testing probabilistic models for use in its official projections[7], although their framework for including uncertainty has yet to be fully defined.

Uncertainty in population projections come from four main sources: the projection model(s), parameter estimates, expert judgments and historical data[8]. Uncertainty can also be based on the results of past projections[9] [10]. In this article we show how historical observations and model assumptions influence uncertainty, as well as the inclusion of expert beliefs regarding future patterns. We do this by applying various autoregressive time series models to population growth rates in England and Wales. Population forecasts are based on past patterns, where a long time series of data are very valuable for assessing our uncertainty for the future.

In nearly all of the probabilistic literature on population forecasting the approach has been from a frequentist (classical) perspective. We introduce a Bayesian approach, which offers population forecasters the most flexibility in terms of specifying uncertainty. Unlike frequentist models, Bayesian models allow for the integration of uncertainty expressed in prior distributions, empirical data and expert judgements. However, these models have yet to be widely applied in the population forecasting literature (see the next section in this article).

As this work is written for a general audience, we have left out the technical details of the models used to produce the Bayesian time series forecasts. For those interested in the specification of these models, refer to Abel *et al*. (2010)[11]. Also, note that this work represents some of the early efforts carried out by a team of researchers in the ESRC Research Centre for Population Change (CPC). In the future we plan to expand these ideas to more complex population models that include, for example, age, sex and state transitions that a population experiences (for example, residential, marriage, and employment).

In terms of structure, we first provide a review of standard population projection approaches and describe the current approach of ONS. This is followed by a section outlining the Bayesian approach to time series forecasting. We then compare our forecasts with official projections by ONS and to alternative forecasts based on a different prior assumption and shortened time series. Finally, we end the article with some conclusions and recommendations for future work.

A review of population projection approaches

Various typologies of macro-level population projection methods can be obtained by applying some simple criteria. In this brief review we focus on three of them: dimensionality of the problem under study (simple extrapolations of population size or growth rates, single region cohort-component models and multiregional models), the approach to uncertainty (deterministic versus stochastic) and methodology (data-driven versus expert-driven). For simplicity we assume that expert-driven methods encompass projections based on theories and expectations about the future. In many cases, however, projections combine aspects of data-driven methods and expert judgements[12]. More detailed typologies can be found in Willekens (1990)[13], de Beer (2000)[14], O'Neill *et al*. (2001)[15], Wilson and Rees (2005)[5], Booth (2006)[16] and Bijak[17].

With respect to the dimensionality of population projections, the simplest models rely on the extrapolations of population size, population growth rates or crude rates related to particular components of demographic change (fertility, mortality and migration). The adding of age and sex leads to the cohort-component framework of population accounting developed by Leslie[18]. Cohort-component models are extendable by adding additional dimensions, such as spatial regions, as suggested by Rogers in his seminal work on multiregional demography[19], or subgroups, such as ethnicity[20]. Here 'multiregional' refers to all multidimensional extensions of the cohort-component model, including other multistate models. See: Land and Rogers[21]; Schoen[22 23] and Rogers[24].

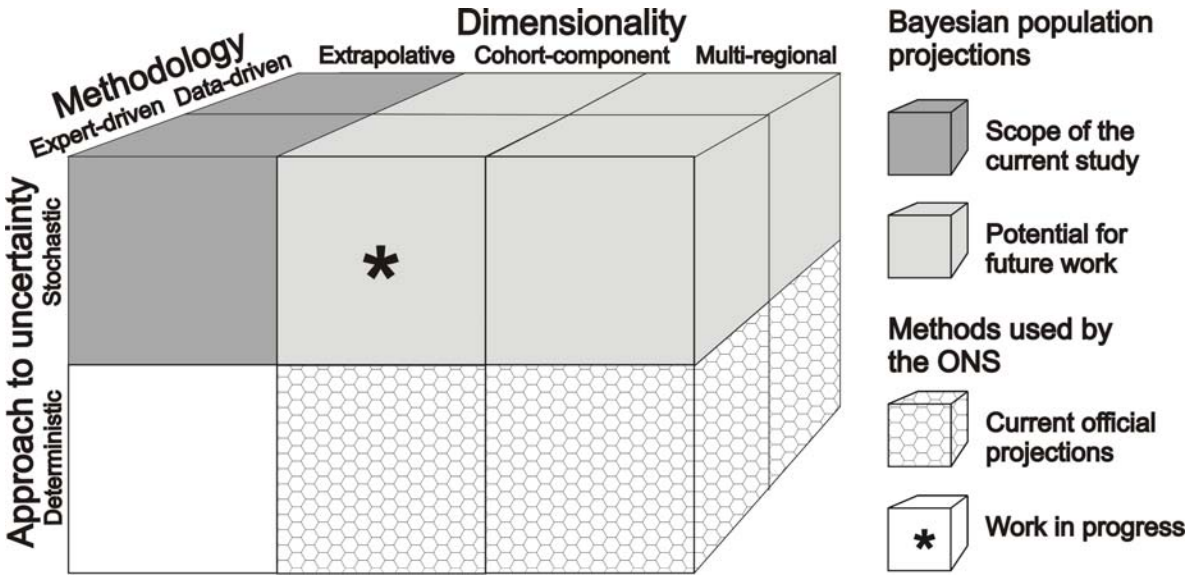
Another feature characterising any method of population projection is the approach to uncertainty. Uncertainty in projections can be ignored, described using various plausible scenarios or quantified using probabilities[14]. The deterministic scenarios can be data-driven, that is, based on simple mathematical extrapolations of past trends, or expert-driven, that is, relying mainly on expert judgement[17]. Similarly, stochastic (probabilistic) projections can be based on time series analysis or extrapolation of past projection errors[8], or based on expert opinion used to assess the future uncertainty[3]. The Bayesian methodology, advocated throughout this article, allows for combining both features in a coherent and consistent way. So far, only a handful of population forecasts have been prepared within the Bayesian framework[25 26].

The current official population projections for England and Wales produced by ONS represent results from a deterministic model with uncertainty, not quantified in terms of probabilities, but denoted by various plausible scenarios[12]. Recently, promising attempts were undertaken to produce expert-based stochastic population projections for the United Kingdom[7]. Both the current and probabilistic work of ONS are indicated in the 'methodology cube' in **Box one** using patterns and an asterisk, respectively.

The philosophy of Bayesian statistics enables the combining of data- and expert-based approaches within a common, stochastic framework. Results are presented in this article for a simple, extrapolative example (darker shading in Box one), but our approach can be extended to

include cohort-components, and eventually multiregional cases (lighter shading). In this way we believe that the Bayesian approach can complement the methodological developments currently undertaken within ONS[7].

Box one Proposed Bayesian methods and the practice of the ONS population projections



Source: Own elaboration.

For more detailed typologies, see Willekens (1990)[13], de Beer (2000)[14], O'Neill et al. (2001)[15], Booth (2006)[16] and Bijak (forthcoming)[17].

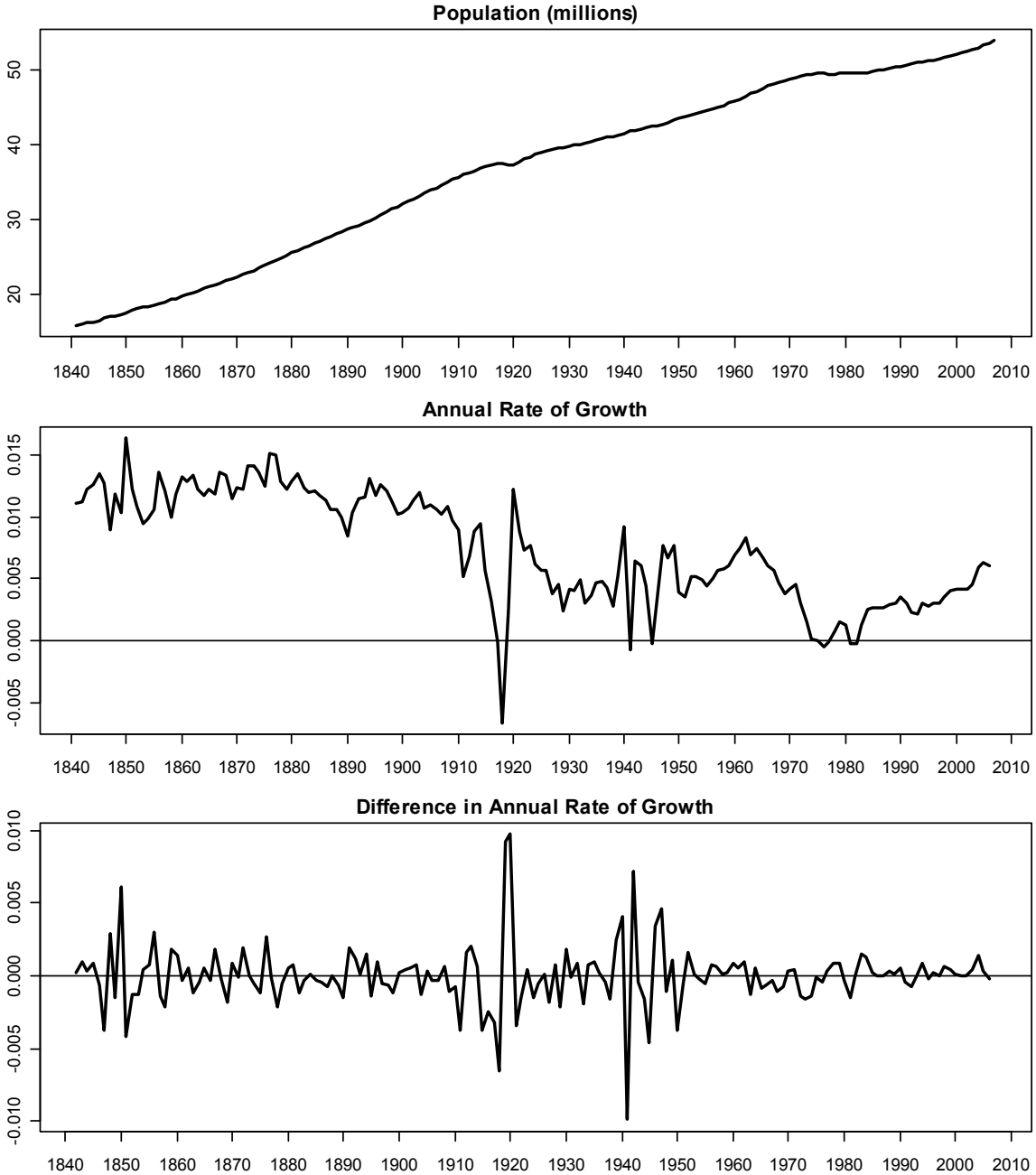
The uncertainty of the future UK population

Bayesian time series models for population forecasting are introduced in this section. First, we present the methodology for Box-Jenkins time series models. This includes transformations to data, autoregressive models for the mean process and stochastic variance models for observed data with a non-constant variance over time. Second, Bayesian methods for estimating parameter values are discussed. Here, the specification of prior distributions and model uncertainty represent the main focus. More detailed information about the models and parameter estimation can be found in Abel *et al*. (2010)[11].

Time series modelling of annual population series

Annual time series of population totals often display some form of trend or fluctuations over long time periods. To illustrate, consider the mid-year population estimates (including military personnel) obtained from the Human Mortality Database[27] for England and Wales from 1841 to 2007, as presented in the top panel of **Figure 1**. The graph clearly illustrates an increasing trend with the population rising from 15.8 million in 1841 to 53.9 million in 2007.

Figure 1 England and Wales population data, 1841-2007



Time series models for population forecasting usually concentrate on the rates of population growth over time, r_t, provided in the second panel of Figure 1. For this work the growth rates are calculated as:

$$r_t = \frac{p_{t+1}}{p_t} - 1 \qquad (1)$$

where p_t is the population total at time t. A standard requirement for fitting time series models is that the data must exhibit (weak) stationarity. This implies that both the mean and the variance of

the data are constant over time. These properties are not present in the historical series of r_t shown in the middle panel of Figure 1. Instead, the series exhibits a downward trend, caused predominantly by falls in mortality and fertility rates from pre-industrial levels.

Experience suggests that, if we are to use time series models which assume stationarity, transformations of the data may be required[28]. One such transformation is to take the differences in r_t, that is:

$$y_t = r_t - r_{t-1} \quad (2)$$

and to model them instead. A plot of y_t is provided in the bottom panel of Figure 1, where a constant mean level, close to zero, is clearly illustrated. The plot also demonstrates peaks during some noticeable historical events, such as the two world wars and the 1918 influenza pandemic, which had dramatic effects on the change in the annual rate of growth. These events may lead to the conclusion that, although the series of y_t has a constant mean, it cannot be considered to be completely stationary as the variance appears non-constant over time. Models to account for this feature are outlined later in this section.

Autoregressive (AR) models have a long history of being used to forecast populations. See for example, Saboia, 1974[29]; Ahlburg, 1987[30]; Pflaumer, 1992[31]; Alho and Spencer, 2005[8]. The key feature of AR models is the inclusion of parameters for the regression of variables such as y_t, on previous values of itself, y_{t-j}, where j represents the time lag. This is commonly known as *autocorrelation*. AR models can include multiple parameters for autoregression at different time lags. For example, an AR model of order 3 is denoted as AR(3) and has autoregressive terms at lag 1, 2 and 3. Time series models also tend to have a parameter for the mean level of the process, represented by μ.

Once fitted, AR models can be used to forecast future values of the time series process. If the process considered is the change in population growth rates, y_t, (as in this article), future values of the original population growth rates, r_t , can be derived by re-arranging Equation (2). In our case the last observed population total is p_{2007} (Figure 1). Based on these data, we can derive a series of population growth rates up to r_{2006} and changes in population growth rates up to y_{2006}. Thus the first step-ahead forecast from an AR model, y_{2007}, can then be used to obtain

$$r_{2007} = y_{2007} + r_{2006} \,. \quad (3)$$

From this, we can derive the forecast of p_{2007} by re-arranging Equation (1) as

$$p_{2008} = (1 + r_{2007})\, p_{2007} \,. \quad (4)$$

Subsequent values of r_t and p_t may be calculated in the same manner, using the forecasted y_t estimated from the model.

As noted previously, historical time series of demographic data often exhibit some volatility due to events such as epidemics, wars or baby booms. This is certainly true for the data presented in Figure 1. Stochastic Volatility (SV) models allow for a non-constant variance when modelling time

series data. This is done by specifying a time-dependent model for the variance, as well as the mean. Consequently, SV models can account for heterogeneity found in the demographic data, allowing forecasts to be adjusted according to the level of volatility estimated at the time the projection is made.

Bayesian time series methods

The estimation of parameters in time series models can be undertaken using a number of different methodologies. In this article we use a Bayesian methodology because both expert opinion and uncertainty in model choice can be included. See **Box two** for an introduction to Bayesian inference.

The incorporation of expert opinion has become an increasingly important input into the prediction of future populations[7]. Bayesian methods allow these opinions to be fully incorporated into the estimation procedure by specifying prior distributions in relation to the model parameters. The distributions can be set to 'flat' if the expert does not have any notions about what the parameter values should be. This results in parameter estimates that are very similar to those fitted by using classical statistical methods. On the other hand, if the expert does have some ideas about what particular parameter values should be, then that person can specify a distribution centred on these values and incorporate them directly into the estimation procedure. The result is parameter estimates that reflect the combination of the expert's prior beliefs and the empirical data.

Bayesian methods allow uncertainties in model choice to be incorporated using probability distributions representing the likelihood for each model. This allows models to be averaged across a set of plausible models, rather than selected, as is the common practice in classical statistics. This is advantageous as it is unrealistic to be sure that any particular model is the right one on which to base forecasts. In addition, Bayesian model averaging can operate across models that are non-nested, such as between AR models and SV models. Finally, Bayesian methods allow the incorporation of model uncertainty to be directly integrated with parameter uncertainty. The result is probabilistic forecasts that are likely to be more realistic than those produced with classical methods.

Box two Bayesian inference in a nutshell

The Bayesian approach to statistical inference dates back to the seminal work of an English nonconformist clergyman, Rev. Thomas Bayes (1763)[32]. The essence of Bayesian inference consists of updating *prior distributions* about the model parameters, θ, in the light of some empirical data, x. The combination of the two results in a *posterior distribution*.

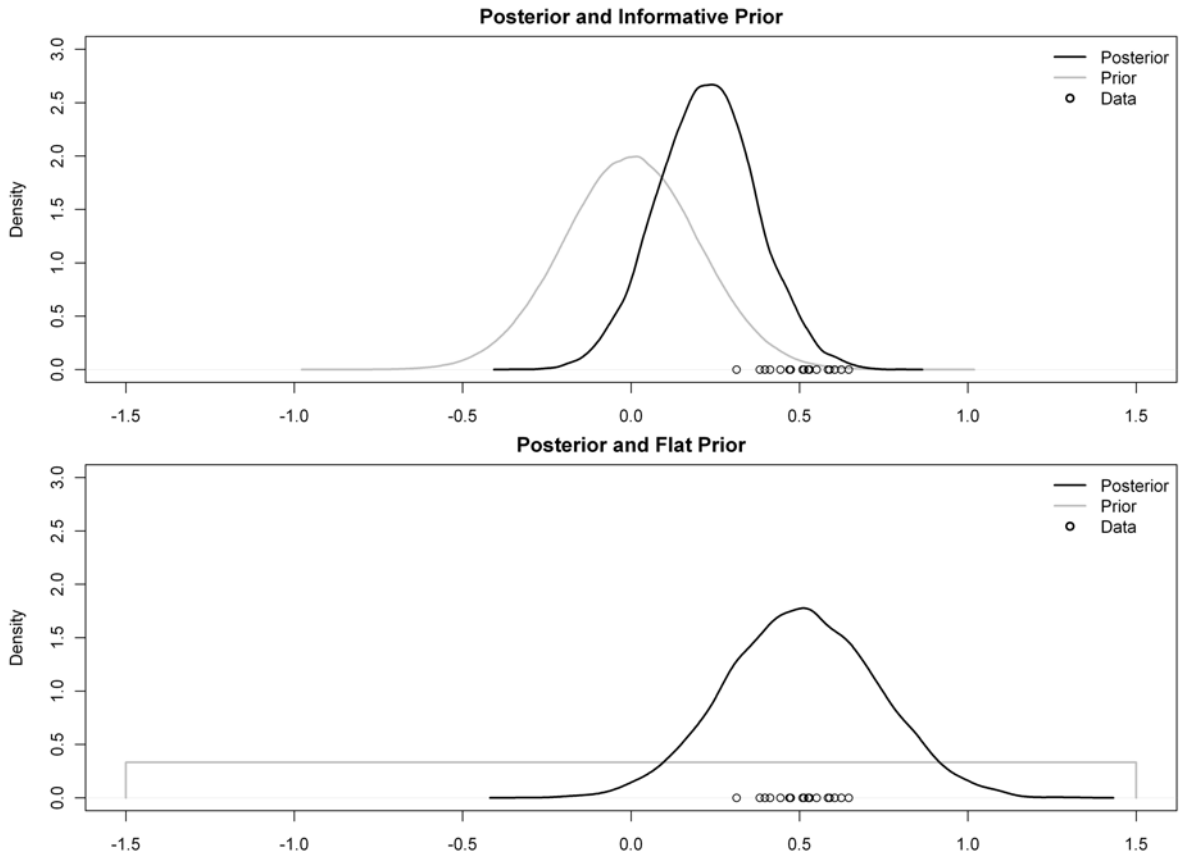
The prior distributions reflect the knowledge or belief of a researcher in different values of θ, without taking the data into account. The prior distributions can be either informative, or rather vague, as it is the case for 'flat' distributions (see the stylised example below). Formally, the Bayes theorem can be written using probability distributions, as:

$$p(\theta \mid x) = \frac{p(\theta) \cdot p(x \mid \theta)}{p(x)}.$$

Box two Bayesian inference in a nutshell

Here, $p(\theta \mid x)$ is the posterior distribution, $p(\theta)$ is the prior, and $p(x \mid \theta)$ denotes the likelihood of data.

Example: In order to illustrate the effect of alternative prior assumptions we simulated 20 observations from a Normal distribution with mean 0.5 and standard deviation of 0.1. We then re-estimated the mean using two alternative prior assumptions. In the first case, a normal distribution with mean 0 and standard deviation of 0.2 was used. The resulting posterior distribution is shown in the top panel of the graph below. In the second case, a Uniform distribution with a lower bound of -1.5 and an upper bound of 1.5 was assumed. The resulting posterior distribution is shown in the bottom panel of the plot. The two plots demonstrate how a posterior distribution can become narrower and alter in central tendency when an informative prior is included.



The combining of prior information and empirical data can be of great benefit when forecasting populations. For example, the future mean in a time series model based on past data may suggest an annual growth rate of 0.5 percent. This may be different from that expected by demographic experts (who may for example expect a future mean annual growth rate of zero). Hence, the inclusion of their opinions as an informative prior can help direct the parameter estimate of the mean level, on which model forecasts are based, away from an estimate that is based on the data and with an uninformative flat prior.

Recently, computation of a Bayesian model has become relatively easier as computational power has become more readily available, including introduction of the WinBUGS software. The latter has allowed users to estimate posterior distributions easily and quickly, without having to programme complex Markov chain Monte Carlo (MCMC) routines. For example, only a few lines of code are required to set up an AR model and state the prior distributions of the model parameters. Posterior distributions of parameters from a converged sample of an MCMC chain can be obtained speedily on a standard desktop computer. One such application of these advances has been in the estimation of parameters in SV models, which typically use Bayesian methods. See for example, Meyer and Yu, 2000[33]; Congdon, 2001[34]; Jacquier, 2003[35]. Bayesian estimation is typically preferred over classical parameter estimation because of the intractable form of the likelihood function[32].

Comparisons of forecasts

In this section results of forecasts from Bayesian time series models fitted to the historical data in Figure 1 are presented and compared with several official population projections. First, we compare our model averaged forecasts to the latest ONS scenario-based projections. We then revise our model averaged forecasts by adding expert opinion for a single parameter to understand better the effect of changing from a flat prior distribution to an informative one (see Box two). In the last section we compare our Bayesian time series forecasts on several shortened data series against past official projections, and to the actual observations.

Model averaged forecasts

For the Bayesian time series forecasts we consider 18 models for the differenced population growth rate, y_t, which are the same as those described in Abel *et al.*[11]. These consist of an independent normal (IN) model (with just a mean parameter and no autoregressive terms) and eight AR models (with non-zero means) that increase in order from AR(1) to AR(8). Nine more models with additional terms to control for stochastic volatility in y_t were also considered. This range of models was selected in order to represent all possible autoregressive processes that might adequately describe the differences in the overall growth rate series. As we had no previous knowledge about the nature of the parameters in each model, we assigned non-informative prior distributions. Also, note that the priors used for each model were the same.

The posterior distributions of the Bayesian time series forecasts can be summarised in a number of ways. In this section, we focus on summaries of the posterior distributions at two levels: the model probabilities and the joint predicted posterior distributions for future values of r_t and p_t. The posterior model probabilities of the 18 Bayesian time series forecasts fitted to several series of y_t with different end-points (see below under 'In sample forecasts') are provided in **Table 1**. The two last columns refer to two forecasts relying on the entire data set, that is, 1841-2007: one with 'vague' (flat) prior assumptions and the other with information assumed *a priori* (see the next section). In both cases the results indicate strong support, with a model probability of between 0.75 and 0.80, for the independent normal with stochastic volatility term (IN-SV). This model has only a single term for the mean level of change in the population growth rates (with no autoregressive terms) alongside parameters to control for the volatility shown in the data. The next most likely model is the AR(3)-SV model, followed by the AR(1)-SV. These models indicate that there is a small degree of support for models that include terms for autoregression at lags 3 or 1. The SV

models with higher order AR terms, in addition to the models with constant variance terms, have very low model probabilities under 0.01.

Table 1 Posterior model probabilities for 18 models fitted for data series with different end points

Model	Posterior model probabilities						
	1957	1967	1977	1987	1997	2007	
						Flat prior	Informative prior
IN	0.00054	0	0	0	0	0	0
AR(1)	0.00021	0	0	0	0	0	0
AR(2)	0.00067	0	0	0	0	0	0
AR(3)	0.00147	0.00001	0	0	0	0	0
AR(4)	0.00051	0	0	0	0	0	0
AR(5)	0.00036	0	0	0	0	0	0
AR(6)	0.00006	0	0	0	0	0	0
AR(7)	0.00001	0	0	0	0	0	0
AR(8)	0	0	0	0	0	0	0
IN-SV	0.29967	0.49045	0.74872	0.79542	0.67155	0.79833	0.74962
AR(1)–SV	0.23621	0.18083	0.18004	0.11968	0.12038	0.07126	0.05314
AR(2)–SV	0.03650	0.03367	0.01731	0.01656	0.03123	0.01762	0.01431
AR(3)–SV	0.39972	0.27773	0.05032	0.06229	0.16113	0.10025	0.16832
AR(4)–SV	0.02240	0.01570	0.00321	0.00551	0.01436	0.01127	0.01332
AR(5)–SV	0.00152	0.00148	0.00038	0.00049	0.00123	0.00117	0.00118
AR(6)–SV	0.00014	0.00011	0.00003	0.00003	0.00011	0.00008	0.00011
AR(7)–SV	0.00001	0	0	0	0.00001	0.00001	0.00001
AR(8)–SV	0	0	0	0	0	0	0

Given the posterior model probabilities from all 18 models, the joint predictive posterior distribution for future y_t up to 2032 was estimated. This provided a sample of 10,000 observations of future y_t values. These were then transformed to obtain the joint predicted distributions of future r_t and p_t using Equations (3) and (4), updated for each subsequent year.

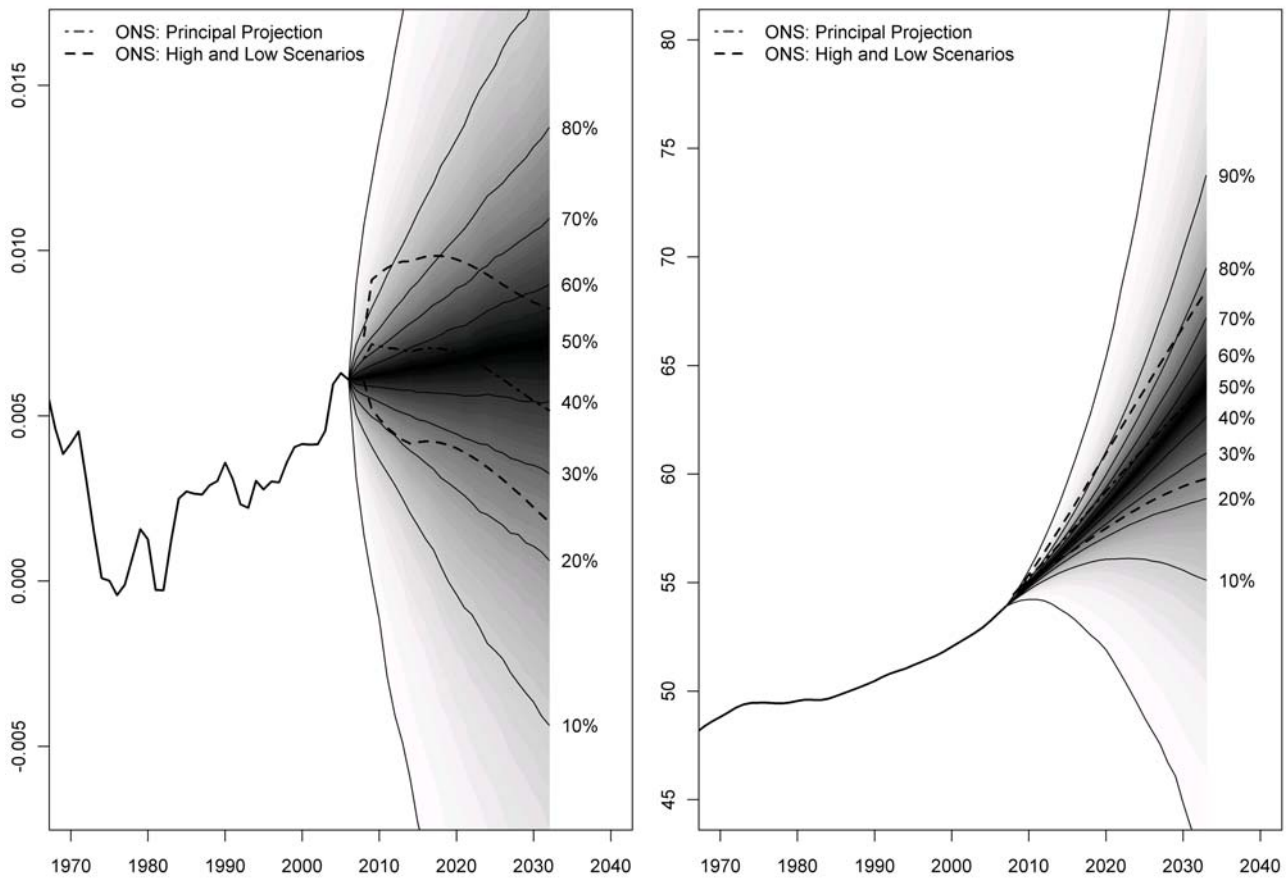
The results are presented in the left and right panels of **Figure 2**, respectively. Each shade of the forecasted fan in these plots represents a single percentile of the estimated posterior density, where the darkest shades correspond to most central values and the lighter shades to the tails of the distribution. Contour lines are also plotted at each decile and at the 1st and 99th percentiles.

From these forecasts the median predictive population in 2033 was 64.0 million. Numerous measures of uncertainty are also available. For example, in 2033 the 20th percentile is 59.0 million persons and the 80th percentile is 69.4 million persons. In other words, our forecasts predict a 60 per cent probability that the 2033 population will fall between these two numbers.

Summaries of the predictive probability distributions can be compared with national projections. In the United Kingdom, ONS regularly prepares a set of projected total populations based on cohort component methodology under a range of deterministic scenarios. For this study we compare our results with approximations for three variants (principal projection, high and low population) published in the latest set of projections for England and Wales[36]. The principal variant relies on

assumptions considered to reflect best the demographic patterns at the time they were adopted. The high (or low) population variant assumes a combination of high (or low) fertility, high (or low) life expectancy and high (or low) net migration, and is intended to provide users with a better sense of the plausible future extremes in population change. All three variants of population totals are displayed on the right hand panel in Figure 2. In the left panel, the derived values of r_t, calculated using Equation (1), and the future values of p_t are shown. The central, dot-dashed lines represent the principal projections, while the upper and lower dashed lines represent the high and low population variants, respectively.

Figure 2 Joint predictive probability distribution of the model averaged growth rates (left) and resulting population forecast in millions (right)



The panels in Figure 2 illustrate a number of differences between the ONS principal projection and that of our model averaged forecasts. First, the uncertainty in the ONS rate, represented by their high and low variants, is narrower than our model averaged forecasts at all points of time. In other words, the Bayesian forecasts include a wider range of uncertainty than those produced by ONS. Second, the uncertainty in the rate of population growth of the ONS projection does not increase substantially over time, unlike those derived using probabilistic methods. The reason for this is that ONS includes the effects of a rapidly ageing population (which affects all mortality variants) in their cohort component projections. Our models, based on overall population growth rates, are unable to

account for these types of effects. Third, the ONS principal population projection in 2033 of 64.1 million is about the same as our model averaged median (64.0 million). Finally, the high and low variants in the projected population totals by ONS lie within the 77th and 22nd percentiles of the posterior predictive distribution of the 2033 population forecasts. In earlier forecast years the population totals from the high projection scenario are greater than our 80th percentile. On the other hand, the projected population totals from the lower variant never fall below our 20th percentile.

Sensitivity to alternative prior

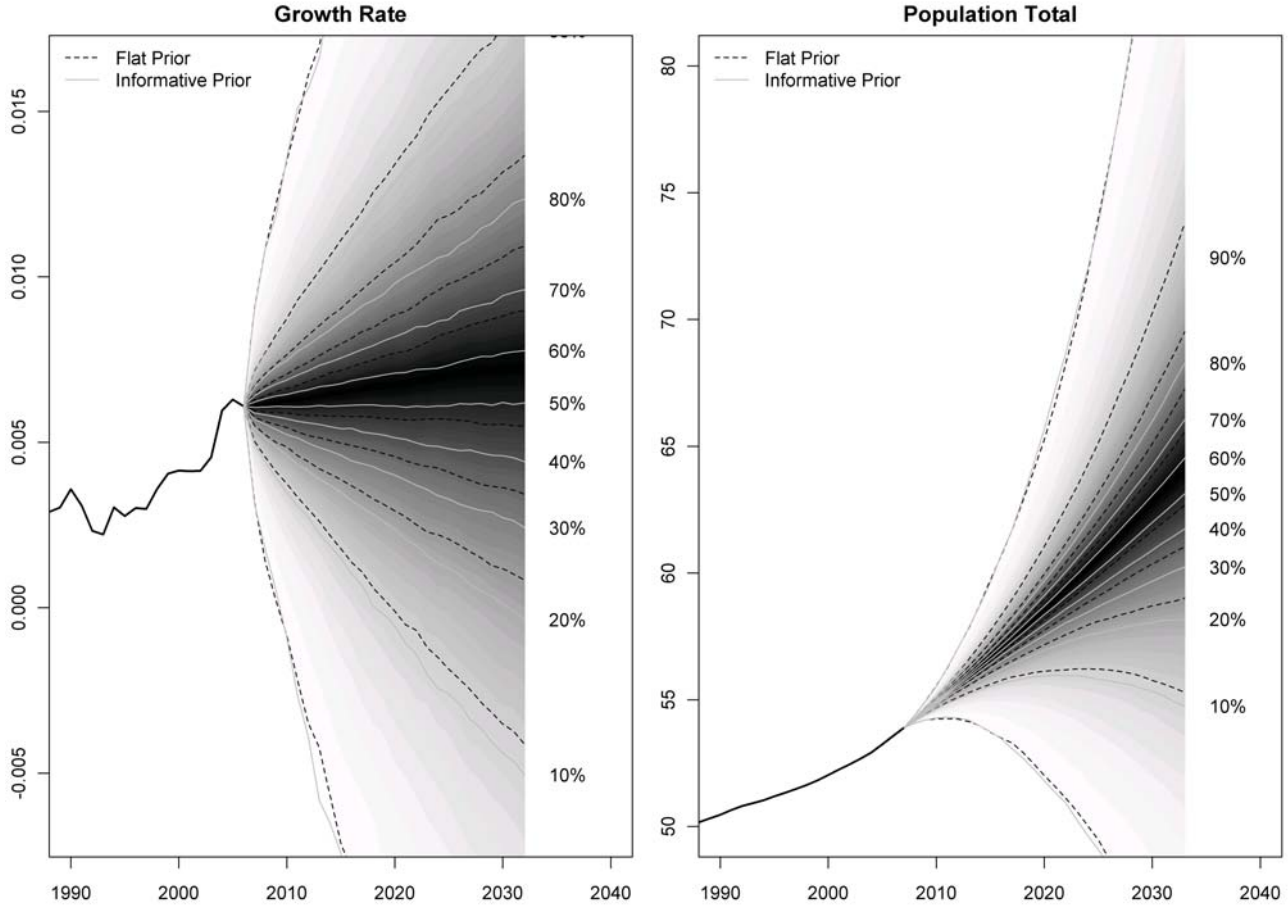
The forecasts presented in the previous section assumed flat prior distributions for all parameters. In this section we analyse the sensitivity of the posterior parameter estimates and model probabilities to the introduction of an informative prior. This is conducted by changing only the mean level of y_t (i.e., μ). This term also represents the annual mean level of increase (or decrease) in r_t and is present in all 18 models. In previous forecasts we assigned a non-informative prior distribution of $\mu \sim N(0, 100)$, where N denotes a Normal (Gaussian) distribution with mean 0 and variance 100.

To establish an informative prior, we used the ONS 2008-based principal, high and low projections to derive each variant's values for r_t and y_t from 2007 to 2032. The mean of the principal projection y_t (-0.000065) was used as the mean of new formative prior distribution. For the variance the means of the high (-0.000177) and low (-0.000035) variants of y_t were assumed to represent the 80th and 20th percentiles, respectively. These were chosen as previous authors, for example, Stoto (1983)[37] and Alho (1992)[38] have previously found high and low scenarios to represent roughly 66% confidence intervals. After a search among candidate distributions we found the $\mu \sim N$ (-0.000065, 0.0001) to approximately meet this criteria.

Given the informative prior, we calculated the corresponding posterior distributions for the parameter estimates and model probabilities. The AR models with informative priors exhibited mean values of μ similar to those with flat priors, albeit with reduced standard deviations. However, in the SV models, the mean values of μ became much closer to zero. The posterior model probabilities for two models with alternative prior assumptions for μ are shown in Table 1. Here we see that the posterior model probabilities of the informative prior remained fairly similar to the models with the flat prior assumptions.

To understand the effects of introducing informative priors on the future population growth rates and population totals, the predictive posterior distributions resulting from both model assumptions are plotted in **Figure 3**. As expected, the two plots on the right illustrate a reduced amount of uncertainty in comparison to the predictive posterior probability distributions obtained from the flat prior assumption (on the left side). For example, the 20th and 80th percentiles of p_{2033} were 58.9 million and 69.3 million, respectively, when the flat prior was used compared to 58.2 million and 68.3 million, respectively, when the informative prior was used. In addition, the median of the predictive posterior probability distribution reduced from 0.00733 for r_{2032} from the flat priors to 0.00619 from the informative priors. Consequently, the median of p_{2033} also falls from 64.0 million to 63.1 million.

Figure 3 **Comparison of predictive posterior probability distributions of the population growth rates (left) and population (right) for flat and informative prior distributions**



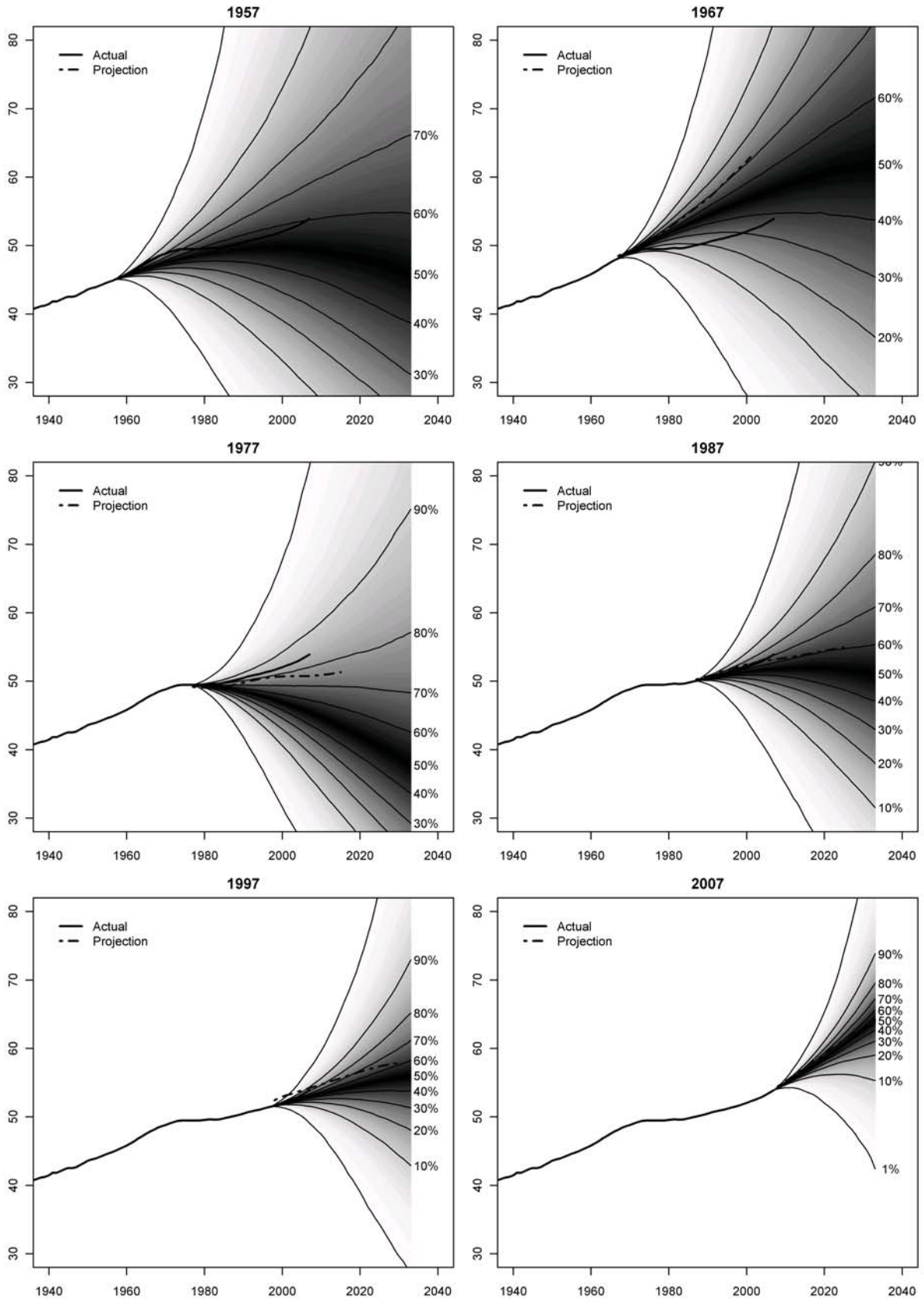
In-sample forecasts

To assess the performance of the Bayesian time series methods, in-sample forecasts (using flat priors) were conducted by using five shortened data sets with end points at 1957, 1967, 1977, 1987 and 1997 respectively. The results of these forecasts are compared against both past official population projections obtained from the Government Actuary Department (GAD) website [39] and with actual observations.

The posterior model probabilities from our forecasts with shortened series are presented in Table 1, alongside the model probabilities for those based on the full length data series ending in 2007. As expected, the forecasts based on longer time series have similar model probabilities as those based on the full data, with large support for the IN-SV model. In the shorter data series forecasts with end points in 1957 and 1967, more support is given to models that include autoregressive terms. This is most notable for the data with the end point in 1957, where the AR(3)-SV model has the highest model probability (0.39972).

The model averaged posterior predictive distributions of p_t for each of the shortened data series, along with the previously presented forecast from full data series (ending in 2007) are shown in **Figure 4**. For the shortest series, with last observation in 1957, the median of the population

Figure 4 **Sequence of joint predictive probability distribution of the population forecasts up to 2033, actual and GAD projections in the past six decades (in millions)**



forecast in the 2033 predictive distribution is 45.8 million. As we move sequentially through the results from data sets of increasing length, the median of the p_{2033} distribution increases to 61.9 million for the 1967 data set, falls to 37.6 million for the 1977 data set, and then increases to 51.0 million to 56.0 million to 64.0 million in the 1987, 1997 and 2007 data sets, respectively. An important factor explaining the large differences between the 1957, 1967 and 1977 forecasts are the 'baby boom' and 'baby bust' periods. In each of these forecasts the historical data did not provide any indications that the direction of population growth would change, thus leading to their relatively low or high projections. This is a weakness of time series forecasts, although it should be noted that experts at the time were also just as confused.

There are a number of noticeable conclusions that can be drawn when comparing the forecasted posterior distributions with the actual data and GAD projections, represented by the solid black line and dot-dashed line, respectively, in Figure 4. The median of the forecasted posterior distributions based on the 1957 data consistently underestimated the actual population (as it could not consider the future increases in fertility). This error was greatest during the early part of the forecast horizon where the actual population strays into the upper tails of our posterior distributions. However, this error improves, especially during the late 1980s when the population total moves towards the centre of our posterior distributions. In essence, the 1957 forecast averaged out the effects of the baby boom during the 1960s and the baby bust in the 1970s. In 2007 the England and Wales population was 53.9 million, which is within the 61st percentile of our p_{2033} posterior distribution. Not surprisingly, the GAD projection of 1957 suffers a very similar pattern of errors as our medians.

The median of our forecasts based on the 1967 data consistently overestimate the population, a consequence of making a forecast at a time of relatively large population growth. The error is greatest in the early part of the forecast where the growth rate of the actual population quickly decreases (see Figure 1). As with the 1957 based projection, the error improves in the latter part of the forecast horizon, with the 2007 observed population lying within the 39th percentile. The GAD projection, made in 1967, overestimates the actual population to a greater extent than our forecast, consistently following the 70th percentiles of our posterior distributions of p_t. This is presumably because they did not rely as much on historical patterns of population change as our forecasts do.

The forecasts based on the 1977 data, during the baby bust, suffer the largest errors of all the in-sample data sets. The actual population consistently remained in the upper tail, between the 80th and 90th percentiles of our posterior distributions, with the 2007 observed population lying within the 85th percentile. This large error is due to a combination of factors. First, the data series for r_t exhibits a turning point in the early 1980s, when the population began to increase once more. Second, unlike previous forecasts for shorter data series, large posterior model probabilities were estimated solely for the IN-SV model. As a result, there is a large reliance of the median forecasts on the μ parameter in this model. In addition, there is a lack of autoregressive parameters to temper the trend effect in the mean process, unlike the 1957 and 1967 based forecasts. The GAD projection in 1971 also underestimated the actual population, but with less error compared to the median of our posterior distributions.

Both the 1987 and 1997 based forecasts underestimate the actual populations, with the 2007 observed population lying within the 71st and 84th percentiles respectively. The reason for the difference is largely due to unanticipated increases in population caused by net immigration. The actual population closely follows the 70th percentile of our 1987 based forecast throughout, while

the GAD projection follows our 60th percentile. The 1997 GAD projection appears to be affected by errors in the intercensal population estimates. As the forecast horizon increases, their projection becomes closer to our medians of the posterior distributions.

Conclusion

In this article we have presented a number of population forecasts for England and Wales. Utilising Bayesian methods, we have introduced uncertainty from multiple sources, including model choice and parameter estimation. We believe the resulting forecasts therefore provide a more realistic summary of future uncertainty in population forecasts in comparison with equivalent time series models fitted using classical methods.

Volatility in population growth rates was controlled for using stochastic volatility models, which tended to have the highest posterior model probabilities when fitted to historical data. The ability to control for volatility may be of importance when considered in the context of cohort component projection methods. These methods often require assumptions about future rates of population growth components. However, previous authors have noted that the success of these assumptions, when comparing their past projections with the actual population, may simply reflect the volatility or stability of the respective time series at the time the projections are made. See Shaw (2007)[40] and Keilman (2007)[41].

Bayesian methods allow the formal incorporation of expert judgement embodied in informative priors, and hence alter the forecasted population characteristics and their levels of uncertainty. The initial forecasts presented in this article were based on hardly informative flat priors and hence resulted in the large level of uncertainty in forecasted population size. This level of uncertainty was reduced through the inclusion of additional prior information. We derived our informative prior from future populations projected by ONS, which were based on expert opinion, on the future rates of the components of population change, and on cohort-component methodology. More informative priors based purely on expert opinions regarding the future of population growth rates could have been included. These could, for example, focus on the prior for the mean parameter, as well as the prior distributions for other parameters in the model, such as the degree of autocorrelation or preferences in models (for example, the inclusion of higher weights on SV models). Such prior information would result in further reductions in the estimated uncertainty due to added information in the parameter estimation and model choice procedures.

The simple time series models used to produce our population forecasts provided alternative estimates to those obtained using cohort component methods. When compared with past official population projections the medians from our simple models performed similarly well. In addition, we were able to provide multiple measures of uncertainty. Our models showed a similar degree of susceptibility to turning points, especially when low posterior probabilities were estimated for models with autoregressive terms. This feature might be tempered through the inclusion of expert opinion. For example, we might provide higher prior model weights to those that include autoregressive terms in comparison to the independent normal models.

We focused only on modelling the change in the population growth rate. This has a number of restrictions when interpreting results. For example, we are unable to provide future forecasts for the components of population change or disaggregate future population by age and sex groups.

We hope to explore these areas further in the future using Bayesian methods motivated by the augments provided throughout this article. In addition, further disaggregation of the population growth rate into components is likely to provide more accurate forecasts and further improvements in the estimated levels of uncertainty.

In conclusion, we believe the future of producing population estimates will require more emphasis on specifying uncertainty so that more informed decisions can be made by population planners and policy makers. The use of time series modelling methods allows a large library of statistical and econometric techniques to be applied to meet these demands. The use of the Bayesian approach in fitting these models also allows for further extensions over classical estimation methods, leading to more realistic forecasts and associated uncertainty measures.

Acknowledgements

This research is funded by ESRC Grant number RES-625-28-0001. The Centre for Population Change is a joint initiative between the University of Southampton and a consortium of Scottish Universities in partnership with the Office for National Statistics (ONS) and the General Register Office for Scotland (GROS). The findings, interpretations, and conclusions expressed in this article are entirely those of the author(s) and should not be attributed in any manner to ONS or GROS. The authors would like to thank Peter W.F. Smith, Jonathan J. Forster, Arkadiusz Wiśniowski, and the two anonymous reviewers for their helpful comments and suggestions.

References

1 Ahlburg, DA and Land, KC, eds. (1992) 'Population forecasting: Guest Editors' introduction'. *International Journal of Forecasting* 8:289-299.

2 Lee, RD and Tuljapurkar, S (1994) 'Stochastic population projections for the United States: Beyond high, medium and low'. *Journal of the American Statistical Association* 89 (428) 1175-1189.

3 Lutz, W, ed. (1996) *The future population of the world: What can we assume today?* London: Earthscan.

4 Bongaarts, J and Bulatao, RA, eds. (2000) *Beyond six billion: Forecasting the world's population*. Washington DC: National Academy Press.

5 Wilson, T and Rees, P (2005) 'Recent developments in population projection methodology: A review'. *Population, Space and Place* 11:337-360 (see page 342).

6 Lutz, W and Goldstein, JR (2004) 'Introduction: How to deal with uncertainty in population forecasting?' *International Statistical Review* 72(1) 1-4.

7 Rowan, S and Wright, E (2010) 'Developing stochastic population forecasts for the United Kingdom: Progress report and plans for future work'. Paper presented at the Joint Eurostat/ UNECE Work Session on Demographic Projections, Lisbon. An earlier version of this paper is also available at: www.ons.gov.uk/about-statistics/methodology-and-quality/imps/updates-reports/progress-report-on-developing-stochastic-population-forecasts-for-the-uk.pdf

8 Alho, JM and Spencer, BD (2005) *Statistical demography and forecasting*. New York: Springer. (See pages 238-240).

9 Keilman, N (2001) 'Data quality and accuracy of United Nations population projections, 1950–95'. *Population Studies* 55:149-164.

10 Keilman, N (2008) 'European demographic forecasts have not become more accurate over the past 25 years'. *Population and Development Review* 34 (1) 137-153.

11 Abel, GJ, Bijak, J, Forster, JJ, Raymer, J and Smith, PWF (2010) 'What do Bayesian methods offer population forecasters?' Working Paper 6/2010, ESRC Research Centre for Population Change, University of Southampton.

12 Office for National Statistics (2009) 2008-based national population projections. Available at: www.statistics.gov.uk/StatBase/Product.asp?vlnk=8519

13 Willekens, F (1990) *Demographic forecasting; State-of-the-art and research needs*. In Hazeu, CA and Frinking, GAB, eds. 'Emerging issues in demographic research', pp 9-66. Amsterdam: Elsevier.

14 De Beer, J (2000) *Dealing with uncertainty in population forecasting*. Department of Population Statistics, Central Bureau of Statistics (CBS), Voorburg.

15 O'Neill, BC, Balk, D, Brickman, M and Ezra, M (2001) 'A guide to global population projections'. *Demographic Research* 4 (8) 203–288.

16 Booth, H (2006). 'Demographic forecasting: 1980 to 2005 in review'. *International Journal of Forecasting*, 22 (3) 547–581.

17 Bijak, J (forthcoming) *Forecasting international migration in Europe: A Bayesian view*. Dordrecht: Springer.

18 Leslie, PH (1945) 'On the use of matrices in certain population mathematics'. *Biometrika* 33 (3) 183–212.

19 Rogers, A (1975) *Introduction to multiregional mathematical demography*. New York: Wiley.

20 Rees, P (2008) *What happens when international migrants settle? Projections of ethnic groups in United Kingdom regions*. In Raymer, J and Willekens, F 'International migration in Europe: Data, models and estimates', eds., pp 329-358. Chichester: Wiley.

21 Land, KC and Rogers, A, eds. (1982) *Multidimensional mathematical demography*. New York: Academic Press.

22 Schoen, R (1988) *Modeling multigroup populations*. New York: Plenum Press.

23 Schoen, R (2006) *Dynamic population models*. Dordrecht: Springer.

24 Rogers, A (1995) *Multiregional demography: Principles, methods and extensions*. Chichester: Wiley.

25 Daponte, BO, Kadane, JB and Wolfson, LJ (1997) 'Bayesian demography: Projecting the Iraqi Kurdish population, 1977–1990'. *Journal of the American Statistical Association* 92 (440) 1256–1267.

26 Heilig, G, Buettner, T, Li, N, Gerland, P, Alkema, L, Chunn, J and Raftery, AE (2010) 'A stochastic version of the United Nations World Population Prospects: Methodological improvements by using Bayesian fertility and mortality projections'. Paper presented at the Joint Eurostat/ UNECE Work Session on Demographic Projections, Lisbon.

27 Human Mortality Database. Available at: www.mortality.org

28 Chatfield, C (2004) *The analysis of time series: An introduction*. Boca Raton: Chapman & Hall/ CRC. Page 26.

29 Saboia, JLM (1974) 'Modeling and Forecasting Populations by Time Series: The Swedish Case'. *Demography* 11 (3) 483-492.

30 Ahlburg, DA (1987) 'Population Forecasts for South Pacific Nations using Autoregressive Models 1985–2000'. *Journal of Population Research* 4 (2) 157-167.

31 Pflaumer, P (1992) 'Forecasting US population totals with the Box-Jenkins approach'. *International Journal of Forecasting* 8 (3) 329-338.

32 Bayes, T (1763) 'n Essay towards solving a Problem in the Doctrine of Chances' *Philosophical Transactions of the Royal Society* 53: 370–418.

33 Meyer, R and Yu, J (2000) 'BUGS for a Bayesian analysis of stochastic volatility models'. *Econometrics Journal* 3 (2) 198-215.

34 Congdon, P (2001) *Bayesian Statistical Modelling*, Chichester: Wiley.

35 Jacquier, E, Polson, NG, Rossi, PE (2003) 'Bayesian analysis of stochastic volatility models with fat-tails and correlated errors'. *Journal of Econometrics* 122(1) 185-212.

36 Wright, E (2010) '2008-based national population projections for the United Kingdom and constituent countries'. *Population Trends* 139: 91-114. ONS does not prepare forecasts for England and Wales jointly, and thus our aggregation of the two countries does not take into account internal migration between them or its uncertainty. Our aggregates for England and Wales have to be read with this in mind and may differ from the official published figures.

37 Stoto, MA (1983) 'The accuracy of population projections'. *Journal of the American Statistical Association* 78:13-20.

38 Alho, JM (1992) 'The magnitude of error due to different vital processes in population forecasts'. *International Journal of Forecasting* 8: 301-314.

39 See: www.gad.gov.uk

40 Shaw, C (2007) 'Fifty years of United Kingdom national population projections: how accurate have they been?' *Population Trends* 128: 8-23.

41 Keilman, N (2007) 'UK national population projections in perspective: How successful compared to those in other European countries?' *Population Trends* 129: 20-30.

A cross country review of the validation and/or adjustment of census data

Rebecca Newell and Steve Smallwood

Office for National Statistics

Abstract

This article reviews existing procedures employed by various countries in the evaluation of, and/or adjustment, either of census data, or of population estimates based upon census data. The work was carried out to ensure all potential demographic techniques are considered by the ONS for the post census evaluation process of the 2011 Census.

Contents

Introduction

The purpose of this article is to review existing procedures employed by various countries in the evaluation of, and/or adjustment either of census data or population estimates based upon Census data. The work was carried out to ensure all potential techniques are considered by the ONS for the post census evaluation process of the 2011 Census.

The focus of the work was to consider in particular what demographic techniques were used when considering census results, and whether changes were made either directly to the census database, or as subsequent additions to aggregate results, for the purposes of estimating populations by age and sex. Several procedures have been developed by individual countries over the past few decades and play an important role in understanding and evaluating the coverage of census data. One of the major problems, however, is that it is very difficult to know when to make an adjustment and what the level of adjustment should be. If you had that information you would be likely to have high quality information about the age and sex of the population and it would probably not be necessary to take a census.

There are a number of different methods used to analyse and adjust census data. Some of these can clearly be classed as statistical, such as using 'capture/recapture' techniques or direct macro or micro comparisons with other data sources, while some are more demographic. Demographic analyses are methods that allow us to measure the dimensions and dynamics of populations such as size, composition and spatial dynamics, and how these features change over time. Such indicators are often derived from administrative records of events like births, deaths, marriages, divorces, diseases and employment. The underlying principle behind using demographic methods to validate census data is that demographic analysis can provide measures, which are independent of the census population counts. This can be used as evidence of how many people were missed on census day and may provide a way of making appropriate revisions to population counts where major discrepancies are found.

The use of demographic analysis in census validation is commonly cited in the literature as a 'possible alternative' or 'independent measure' of population counts at the national level (macro level analysis), but most papers do not provide any further information on the actual methods postulated. A review by Don Kerr[1] during the 1990s presented the various procedures used in estimating net undercount of censuses (prior to the millennium round of censuses) in Canada, the USA, Britain and Australia. Data collection methods were briefly described; micro-level procedures (record linkage) outlined; and macro-level procedures (demographic analysis), examined and found to vary in importance and application between countries. Relevant methodologies are discussed below and under the section headed 'Post Census Adjustment Using Demographic Methods'. However, since Kerr's review, little research on the explicit use of demographic techniques for evaluating census data is currently available, and even more sporadic are published examples of the application of demographic analysis in validating or adjusting census figures. From the published literature, there are a number of inter-related techniques that predominantly focus on constructing population cohorts.

Approaches used in demographic analysis

Cohort style approach

A cohort style approach is where estimations are made of expected numbers of the population from known information on births and mortality rates (and in some cases known migration). This works on the simple principle that a known population aged *x* will be aged *x*+*t*, in *t* years. These analyses can be used to examine internal consistency of population counts in terms of their conformity with expected age, period and cohort patterns of coverage; examples are summarised below.

Charlton *et al*. (1997)[2] stated that, 'cohort analysis is a way of providing estimated population counts, independent of all censuses'. They traced cohorts based on Government Actuary life tables at the national level (England and Wales) between Censuses 1971, 1981 and 1991. Results from the cohort analysis provided estimated population counts most similar to the official population counts for 1981, and concluded that 1981 provided the best base from which to roll forward estimated population counts for the 2001 Census benchmark. The benefits of utilising this type of comparative methodology were further illustrated within their analysis, which illustrated the problem of rolling forward the undercount of the 1960s inflow of migrants, where more people were found in the census than the cohort analyses predictions for ages 30-49 in 1971, 40-59 in 1981, and 50-69 in 1991. This approach also clearly demonstrated the emerging phenomenon of 'missing young men' from Census 1981 onwards.

Takami (2003)[3] addressed a similar issue of the 228,561 'age-unknown' persons to appear in the Japanese 2000 Census (i.e. those who were counted in the census but 'age' was missing). An age restricted cohort analysis was used to compare the results of the 2000 Census and the previous 1995 Census. The cohort consisted of persons under 50 years of age, as the number of deaths was believed non-negligible in five year age groups beyond 50. The analysis showed discrepancies between the expected reduction in age-specific populations due to deaths and moves abroad, and the numbers actually missing for the 2000 Census age groups. They concluded that 'age-unknown' persons were most likely to be concentrated around the 20-29 year age group, as the number missing from this age group in the 2000 Census compared to the expected reduction due to deaths and moves abroad, was greater than for all other age groups. Although evidence was found of age-related discrepancies, no formal adjustments were made.

An example of more recent work relates to the quality assurance project carried out by ONS for the 2001 One Number Census (ONC)[4]. This involved comparison of ONC counts against diagnostic ranges derived from rolled-forward population estimates (as advised by Charlton *et al.*, 1997[2]) and the best available aggregated administrative data sources (such as birth registration and pensions data). Where differences between ONC estimated counts and diagnostic ranges existed, extensive checks of the ONC results and diagnostics were undertaken. However, as the authors stated, 'it was never the intention for the census to be adjusted to these diagnostics' and as discussed later, a decision subsequently questioned by academics (see 'Use of Additional Sources').

Demographic age-sex ratios

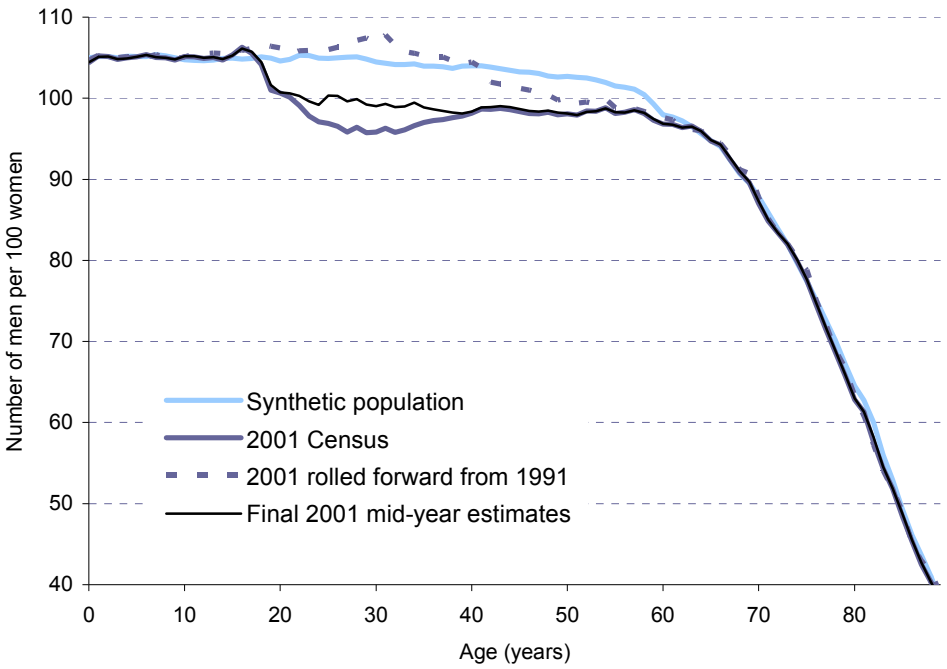
A second body of inter-related literature refers to comparisons with demographic age-sex ratios. These are predominantly derived from the cohort analysis described above. These provide auxiliary evidence for the growing phenomenon of 'missing men'.

Sex ratios will be affected by *relative* undercount, or overcount, of males and females. One particular piece of analysis that can be carried out is to construct from births data and mortality rates the numbers expected to be alive. The overall result of these calculations allows age/sex ratios to be calculated in the absence of migration. For example, work presented at the European Association for Population Studies (EAPS) conference in Barcelona[5] demonstrated for England and Wales a dipping of the male to female sex ratio below 100 from age 20 onwards, as opposed to the natural reversal in the sex ratio from approximately age 50 onwards. **Figure 1** shows the UK Census 2001 age sex ratio against a synthetic population from historical cohort data. The synthetic population was derived from the population of England and Wales born estimated to be alive today, and created by applying historical age-specific mortality rates from Government Actuary Department (GAD) to births.

Further the population figures created can then be compared with the numbers found alive in the Census who were born in the country. This number should always be lower than the estimate of those expected to be alive. It can be inferred that the difference are those that are alive outside the country. The sex ratio for this group can also be computed. However, this latter calculation is not completely independent of the Census as Census data is used to infer the sex ratio of the England and Wales born population abroad. Rather it raises the issue of the plausibility of the resulting ratios.

General findings from around the world relating to this type of analysis[6] elicit similar patterns between census data and demographic age-sex patterns to those depicted in Figure 1. When data is further disaggregated, by ethnicity for example[7] [8] the discrepancy between sex ratios at younger ages is even more evident.

Figure 1 Sex ratio patterns for England and Wales in 2001 from different estimation methods



Further examples of the use of demographic analysis

The US Census Bureau have used demographic analysis since 1960 to assess coverage levels of the census by providing estimated population counts, independent of census population counts by age, sex and ethnicity. These demographic estimates are derived from a culmination of births, deaths, legal immigration and emigration figures, supplemented by data on armed forces, and estimates of undocumented immigration. However, due to the lack of completeness of birth registration data prior to 1940 in the USA, administrative Medicare enrolments are used to estimate the population aged 65 and over.

Wolter (1990)[9] and Bell (1993)[10] explored ways to combine results from US demographic analysis with estimates of census undercount (derived from Dual System Estimation techniques, which are discussed later in this article). Specifically, alternative strategies of incorporating information on sex ratios were investigated, but the findings were complex and were not adopted by the US Census Bureau for the 2000 Census. This is covered in more detail under the 'Uncertainty Models' section below.

Post Census 2000, Robinson e*t al*. (2002)[7] provided a detailed analysis of the differences between the US Census 2000 counts and initial demographic analysis by age and sex. The results showed a clear undercount in the census figures, but that the under-coverage was disproportionate by age and sex, with higher numbers in the demographic estimates for both males and females in the age group 18-29. The findings were put forward as a discussion point for future assessments of the quality of demographic analysis estimates within the US.

Bycroft (2006)[11] explored data quality issues with New Zealand census population counts and those derived from demographic analysis, using sex ratios as their point of comparison. Census counts in New Zealand have shown steady decline in sex ratios since the 1960s, but this decline is not reflected in demographic population counts derived from changes in births, deaths and migration, raising questions over the reliability of more recent census disaggregation by age and sex.

Of the research papers and statistical institutions that indicated discrepancies between demographic analysis and census data, none have led to any formal adjustments to national census figures based on the findings from demographic analysis.

Post census adjustment using demographic methods

The majority of census taking countries in the developed world produce at least two sets of figures: initial census counts that are released soon after or within two years of the census, and some form of 'estimated resident population count', which is separate from the census counts. It is within these estimated population figures that certain countries have adopted some degree of demographic adjustment. However, information on this is limited but key points are highlighted below.

The Australian Bureau of Statistics (ABS) uses 'demographic adjustments' at the sub-national level (Australian States) by age and sex within the preliminary process of producing their 'initial' Estimated Resident Population count (ERP). The demographic adjustments are derived from data on demographic events compiled since 1921 (births, death and overseas migrations), using conventional cohort component analysis, resulting in estimates for the population aged 0-69.

A further comparison is made with the sex ratio in each age group of people enrolled for Medicare, which they believe to be especially useful at the very youngest and older ages where accuracy of the Medicare data is deemed most reliable.

A combination of weighted results for the Post Enumeration Survey (PES) and Medicare data are used to adjust the age groups 70 years and over, whereas a weighted comparison of all three methods (PES, demographic analysis and Medicare data) are used for the population aged 0-69. Unfortunately, the criteria or confidence intervals upon which they base their decision to make direct changes are not stated in the technical papers provided[12 13].

Australia later produced 'final' ERPs that included the addition of Residents Temporarily Overseas (RTOs) on census night (discussed in more detail later). Further sub-state demographic adjustments, which include checks against Electoral Roll data and armed forces adjustments, plus or minus the standard birth, death and migration figures since the elapse of the census date are administered. Essentially ERP undergoes two stages of adjustment, resulting in an 'initial' ERP and a 'final' ERP. The main difference between the two is the amount of information used in each adjustment with the belief that as time goes on, more information becomes available for subsequent adjustments and hence the best possible estimates are produced.

In the production of Population Estimates for Statistics Canada, the initial estimates of net undercount (derived from the Coverage Error Measurement Program), are formally adjusted where implausible discrepancies are found when compared to demographic components of population change (births, deaths, migration and immigration)[14]. An example of where a direct adjustment was made relates to the 1996 Census, where the Reverse Record Check (RRC) results for children aged 0-4 provided an estimated net undercount significantly higher for female children with no obvious explanation. Consequently, the RRC and census results combined gave an implausible figure and subsequently, the number of females was revised to correspond with the demographic age sex ratio[1].

In a similar fashion to Canada, Statistics New Zealand adjusts their 'estimates of net undercount' for ages where large discrepancies exist. For example, adjustments were made for ages 0-9 where substantial discrepancies were found when comparing to the standard demographic components of population change[11], but again, no criteria for the adjustment is cited.

Non-demographic methods of direct census adjustments

In 2001 the United Kingdom developed the One Number Census, which included direct adjustments to raw census count figures, on the basis of Dual System estimation (described below), designed to result in only one set of census population figures being released; the first of its kind[15]. At present the only other country, apart from the United Kingdom, to make any form of 'direct' adjustments to their raw census count figures is Canada, with an overview of the methodology presented below.

Direct adjustments to the census database are administered by Statistics Canada using their post-Census 'Dwelling Classification Study' (DCS), formally known as the 'Vacancy Check'. There are actually four coverage studies used by Statistics Canada (DCS; Reverse Record Check (RRC);

Automated Match Study (AMS); and Collective Dwelling Survey (CDS) but the DSC is the only one that results in the addition of households and persons to the census database.

The DCS adjusts for unoccupied and non-response dwellings (for example classification errors). If a dwelling was occupied in the DSC, one of two adjustments is made to the census database:

1. If the dwelling was listed as 'un-occupied' in the census, then persons deemed missing are added to the census database using the random addition technique (a form of imputation).

2. For non-response dwellings (for example, occupied dwellings for which a completed census was not received), an adjustment is made by creating a new household size for them on the census database.

The results of the remaining three coverage surveys (RRC, AMS and CDS) relate to the calculation of the base population for the post census population estimates as opposed to direct census adjustments[14].

Non-demographic methods of post census adjustments for population estimates

Post Enumeration Survey (PES)

Measured levels of census undercount has increased as a whole over the past few decades, with priority and focus being given to measuring this differential undercount. Consequently, most census taking developed countries now undertake some form of coverage assessment and adjustment, usually in the form of a post-enumeration survey (PES). Detailed examples from each country (Australia[16], Canada[17], Japan[18], New Zealand[19], United Kingdom[20], and USA[21]) can be found in respective NSI technical documents.

Coverage Surveys

The Reverse Record Check is a comprehensive record-linkage method used in Canada by which official gross under and over-coverage is estimated. The system that processes the addresses of the selected persons is automated; however, it is not clear from the technical paper what source these 'addresses' come from. After the sample is taken, persons who should have been enumerated and were not, are then traced and interviewed and the results used to determine the number and characteristics of unenumerated persons. Some limitations to this methodology exist where non-sampling error cannot be quantified in the estimates (that is, those who were untraceable and could not be interviewed). Coverage surveys are also used to identify over-coverage, where people can be counted more than once. The Automated Match Survey (AMS) detects erroneous enumerations by a combination of matching procedures, to identify duplicate households in the census database (for example, persons enumerated in more than one household within the same region). The Collective Dwelling Survey (CDS) estimates over-coverage resulting from persons enumerated as usually resident in non-institutional collective dwellings who are also enumerated at private dwellings. The Reverse Record Check (RRC) was designed to measure over-coverage from all sources (RRC, AMS and CDS). If over-coverage is detected in the AMS or the CDS, then person records are removed from the RRC to eliminate multiple counting of over-coverage between the three coverage surveys. However, from the 2001

Census onwards the CDS component was dropped, and over-coverage estimates for this group (historically very small) were made instead as part of the RRC[14].

Dual System Estimation (DSE)

Dual system estimation is another method that can be used to generate population estimates from census data. The USA ruled out the possibility of using a record-linkage study like the Reverse Record Check (RRC), concluding that such procedures were inappropriate in a country that carries out a decennial census, as tracing records from ten years previously would be too problematic. Instead, the USA use a technique called 'Dual-System Estimation' (DES) where the PES consists of two basic samples: the P sample (area sampling) and the E sample (sample of census enumerations from census database). The proportion of the P sample actually enumerated in the PES provides an estimate of gross under-coverage and is validated through re-listing, re-interviewing, follow-up interviews and record matching. The estimates of under and over-count from P and E samples are then combined to provide a 'dual-system estimate' of net undercount. More detailed information of the procedure can be found in Kerr (1998)[1]. However, DSE is subject to both sampling and non-sampling errors, although continuous efforts are being made to minimise these errors. Alternatives to the dual-system estimates were tested prior to the 2000 Census, in the development of an Integrated Measurement Program, but the attempts were unsuccessful [1 22]. As we have already seen, this methodology can also be used to directly adjust the Census database, for example, in the United Kingdom's 2001 Census[15].

Residents Temporarily Overseas (RTOs)

Both Australia and New Zealand utilise international migration data derived from passenger arrival and departure cards in the twelve months following census. New Zealand's migration data is further supplemented with the capture of electronic passport data during the same 12 month period. In both cases, international migration data is analysed for duration of stay overseas to determine whether persons were missing on census night. Adjustments are then made to Estimated Residential Population counts (EPRs) accordingly[13 23 24].

Due to a change in classification for net overseas migrations (NOM), now defined as being in/out of Australia for 12 out of 16 months, a new two-part method has been proposed for RTO adjustments of future censuses in Australia[23]. At present, New Zealand have not changed their NOM definition but are closely following the outcome of ABS's change in definition and subsequent data output.

Explorative work in New Zealand[11] investigated a new method to improve measures in the change in population due to external migration. At present, no adjustments are made on this basis, but could be incorporated in future censuses. As Castle and Miller (2003)[25] point out:

'.... the experience of migration and of living in another country often leads to modification of the original plans, so that migrants' intentions at the time of departure are poor predictors of actual behaviour'

Bycroft's aim was to use total net migration over a period of time with corrections for incomplete short-term travellers at the start and end of the period, thus adjusting for any 'change' in travel intentions as opposed to taking peoples 'intentions' as definitive. Preliminary results suggest the real contribution of migration to population change is not measured well by current Permanent and

Long Term (PLT) migrant classifications and the new method provides a promising avenue for further research into the effects of misclassification of short and long term migrants.

Use of additional sources in census validation

The United Kingdom

The use of additional data sources in the estimation process has been continually suggested within the literature, but as yet little has been taken forward. Prior to the One Number Census 2001, Chappell *et al.* (1997)[26] stated that administrative sources for the population as a whole and for certain population subgroups needed to be identified along with their quality and availability. Following this, a variety of demographic and administrative record checks (Child benefit; Pension data; MYEs; Home Office prison data; HESA and LSC data) were used to compare the results of Census 2001. No demographic adjustments were made but two of the administrative record checks led to direct adjustments in the ONC process for students' halls of residence figures and DASA data for the armed forces[4]. In summary, Diamond (2003)[4] stated that child benefit and pension data at the time of the 2001 Census were deemed the most accurate and nationally consistent administrative sources available in the UK.

A review of the ONS One Number Census count and quality assurance checks was published in 2007[27]. The collection of independent demographic data and administrative records used with the intention of diagnosing where the census estimates had fallen considerably outside what was expected were discussed. The author questioned why, when over 20 per cent of the 14,000 comparisons made fell outside the more lenient 'demographic and administrative diagnostic range' compared to the 95 per cent confidence intervals for the ONC design, subsequent adjustments were not made to the estimates. The recommendation put forth was that a third source from administrative records *should* be used in future adjustments where discrepancies outside tolerance levels exist. However, no third administrative data source with full population coverage that is more reliable than the census exists.

Looking ahead to the 2011 Census, while quality issues with administrative sources exist, ONS have postulated that the use of additional sources is vital for the 2011 Census and have highlighted some potential sources for use in the estimation process and diagnostic ranges: 'Visitor Data'; the 'Census Household Frame'; and, the use of improved survey data, such as the 'Integrated Household Survey' (IHS) which should be available for 2011[28]. In addition, for the first time in 2011, information on birth and death rates by area will be used to help quality assure census data.

Outside the UK

Both Australia (ABS) and the United States use Medicare data to check the coverage in their population aged over 65 and to make subsequent adjustments to their population estimates where necessary[7 10 12 29]. The ABS also looks at their Electoral Commission (AEC) population distribution against their census population distribution and make further adjustments at the subnational level where clear deviances exist[11].

Japan has investigated Basic Resident Registers (BRR) as a validation check, which cover only those who have Japanese nationality and compare the population census and the BRR with

respect to persons of Japanese nationality. At present no adjustments are made due to questions over the reliability of the BBR data (i.e. non-reporting of moves)[3].

Although Sweden uses a population register as opposed to a traditional census, they have looked at using mortality data to estimate indirectly over-coverage by studying the differences in mortality between immigrants and the native population. The mortality levels for immigrants are compared with those of people born in Sweden after standardising for various background variables. When immigrant mortality in the population registers is very low compared with native Swedish mortality, relative over-coverage in the Resident Total Population (RTP) is assumed. But as the author states, this method has its inadequacies since factors other than errors in the registers may lie behind the differences[30].

Alternative types of estimation

Triple System Estimation

Triple system estimation has been suggested as a method to counterbalance the standard problem of dependence in dual system estimation, where a third list is matched to both Census and PES survey data[27 31]. In both the 1990 and 2000 US censuses, expected sex-ratios were used as the third source in the estimation process, but the data were not formally adjusted in published Census counts[10]. In the UK, it was concluded[31] for the 2001 Census that although some properties are missing, a combined PAF (Royal Mail Postcode Address File) and Census list is currently the best source available, but realistically the resolution is not sufficient enough to act as a third data source at the individual level.

Uncertainty Models

Research during the 1990s explored ways to combine results from demographic analysis with dual-system estimates of net undercount obtained from the PES[9 10]. The original work by Wolter (1990)[9] used sex-ratios to introduce systematic revisions of dual system estimates. Bell (1993)[10] then extended this work by attempting to modify the 1990 PES dual-system estimates using demographic estimates of national sex-ratios by age. The U.S. Census Bureau believes sex ratios to be a robust tool for Census evaluation, and although raw figures are not actually adjusted, post Census figures incorporate these methods[1].

Robinson *et al*. (1993)[7] then developed 'uncertainty models of demographic analysis' that evaluated the uncertainty associated with the measurement of each component of the demographic estimates, including assigning confidence intervals to estimated sex-ratios by age. Their work was subsequently incorporated in the development of US demographic estimates at the subnational level for Census 2000.

Bayesian Methods

The application of Bayesian methods to address uncertainty and 'missingness' in census data, and the framework in which it 'borrows strengths' from additional sources (combining data) is positive in terms of its statistical theory. Bayesian methodology follows three basic principles:

1. Specification of a probability model that includes some prior knowledge about the unknown parameter values (for example, missing persons)

2. Updating of knowledge about unknown parameters by conditioning this probability model on observed data (for example, demographic measures)

3. Evaluate the fit of the model to the data and the sensitivity of the conclusions to the assum ptions.

The few relevant publications that have attempted to explore the use of Bayesian methods in post census population estimates are:

Dick and You (1997a; 1997b)[32 33] attempted to create a more flexible model compared to conventional population estimates where undercoverage rates compared to the national rate in Canada were fixed across provinces. Instead, their model permitted different levels of underlying provincial under-coverage to exist, by allowing the exchangeability of provinces to be examined under a hierarchical Bayesian model[32]. They concluded that some improvement over direct population estimates was possible, but that the methodology was exceeding technical for little, if any quantifiable gain.

Elliot and Little (2000; 2005)[34 35] evaluated a number of models that have been proposed within the USA to deal with post census adjustments, and chose what they believed to be the most robust and placed it within a Bayesian framework. Their Bayesian approach removed the problem of negative cell counts inherent in maximum likelihood methods (MLE), but provided no conclusive outcome as to whether their methods were any better than those already used. The direct benefits of using a Bayesian approach in census validation and or adjustment is currently inconclusive.

Summary and conclusions

Few countries have explicitly used demographic analysis in their evaluation process, and very few use such analysis to adjust either Census counts or population estimates based on the Census. Those which have either used the findings to highlight where possible discrepancies lay within their published census counts (but do not make direct adjustments), or else utilised the demographic discrepancies from the census counts to make post census adjustments for various subpopulations at the national level in their 'Estimated Population Count' procedure.

There are known limitations to the use of demographic analysis: in fact, if demographic analysis was able to give a completely reliable answer to population numbers by age sex and geography, then a census would not be necessary. Aggregate administrative data may contain errors or be incomplete; thus linkage to such data, assuming legal gateways are open, is reliant on the quality of the linkage and interpretation of the unlinked cases. Survey data may not be representative of population characteristics due to non-response, but even if it is representative, will have confidence intervals wider than the accuracy required for Census at low levels of geography. Further, weighted survey data will be weighted to population estimates – the very estimates for which the census is needed. Rates derived from vital events give an indication of potential issues with census populations, but not a precise level for adjustment.

However, the importance of demographic analysis should not be overlooked. Sex-ratios are a robust tool for evaluating age-sex patterns found within the census and should be a consideration in any post census adjustment. One of the major problems relates to confidence intervals around census data, both in terms what level of accuracy is required and whether adjustments actually

improve the accuracy, that is, when does one *really* need to adjust? Unfortunately, limited information was available on this issue from the National Statistical offices discussed.

The usefulness of additional sources of administrative data should also be considered, and utilised where obvious discrepancies are found. Ideally, one would use a third 'total population' data source to both cross reference and help validate census data, but as that 'elusive' third dataset does not exist within the UK a combination of smaller scale comparisons of subpopulation data could be used in an amalgamation of 'small but significant' adjustments where the census data falls short of expected or known profiles. In particular, this must be the evaluation of subgroups of the population where specific data sources for these subpopulations are in most cases fairly accurate (for example, the elderly, the young, school children, students and the armed forces).

References

1 Kerr, D (1998) 'A review of procedures for estimating the net undercount of censuses in Canada, the United States, Britain and Australia'. Statistics Canada, Research Paper, Demographic Documents; Catalogue no. 91F0015MIE-No.005.

2 Charlton, J, Chappell, R and Diamond, I (1997) 'Demographic analysis in support of a one number census'. ONC ONS(SC)97/15, ONS, Census Division.

3 Takami, A (2003) 'Evaluation of the accuracy of the 2000 population census of Japan'. 21st Population Census Conference. Japan, Kyoto.

4 Diamond, I (2003) 'Key Issues in the quality assurance of the one number census'. *Population Trends* 113: 11-19.

5 Smallwood, S and De Broe, S (2008) 'Unravelling sex ratios, what can they tell us about the Census population estimates and recent trends in migration?' Paper presented at the EAPS conference in Barcelona 10-12 July 2008.

6 Smallwood, S and De Broe, S (2009) 'Sex ratios patterns in population estimates'. *Population Trends* 137: 41-50.

7 Robinson, JG, Ahmed, B, Das Gupta, P and Woodrow, KA (1993) 'Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis'. *Journal of the American Statistical Association* 88 (423) 1061-1071.

8 Robinson, JG, West, KK and Adlakha, A (2002) 'Coverage of the population in census 2000: results from demographic analysis'. *Population research and policy review* 21 (1-2) 19-38.

9 Wolter, KM (1990) 'Capture-Recapture estimation in the presence of a known sex-ratio'. *Biometrics* 46 (1) 157-162.

10 Bell, WR (1993) 'Using Information from Demographic Analysis in Post-Enumeration Survey (PES) Estimation'. J*ournal of the American Statistical Association* 88 (423) 1106-1118.

11 Bycroft, C (2006) 'Challenges in Estimating Populations'. *New Zealand Population Review* 32 (2) 21-47.

12 ABS (2006) 'Census Undercount. Demographic Estimates and Projections: Concepts, Sources and Methods'. August 1999. Available at: www.abs.gov.au/ausstats/abs@.nsf/7d12b0f6763c78caca257061001cc588/F66C2B24DC1AE3FFCA25697E0018FC5A?opendocument

13 ABS (2006) 'Rebasing Australia's population estimates using the 2006 census of population and housing'. *Australian Demographic Statistics*, September 2006.

14 Statistics Canada (2004) 'Coverage: 2001 census technical report'. Statistics Canada Catalogue number 92-394-XIE. Available at: www.statcan.ca/english/census01/Products/Reference/tech_rep/coverage/offline%20documents/92-394-XIE.pdf

15 Office for National Statistics (2005) '*Census 2001 review and evaluation: One Number Census Evaluation Report*'. ONS, London.

16 www.abs.gov.au/websitedbs/D3310114.nsf/Home/census?opendocument#frombanner=GT

17 www.statcan.ca/english/census01/P roducts/Reference/tech_rep/coverage/offline%20 documents/92-394-XIE.pdf

18 More detail available at: www.stat.go.jp/english/data/kokusei/index.htm

19 Available at: www.stats.govt.nz/analytical-reports/post-enumeration-survey-2001/chapter-three-post-enumeration-results.htm

20 ONS 2001Census info available at: www.statistics.gov.uk/census2001/pdfs/oncguide.pdf

21 Available at: www.census.gov/dmd/www/refroom.html

22 Mulry, MH and Griffiths, R (1996) Comparison of Census Plus and Dual System Estimation in the 1995 Census test. AMSTAT. Available at: www.amstat.org/Sections/Srms/Proceedings/papers/1996_146.pdf

23 ABS (2009) ABS 2006 Census Rebasing Papers: Australian & State/Territory RTO's: Census Night 2006. *Internal paper; unpublished.*

24 Statistics New Zealand (2009) Post enumeration survey results. Available at: www.stats.govt.nz/analytical-reports/post-enumeration-survey-2001/chapter-three-post-enumeration-results.htm

25 Castles, S and Miller, MJ (2003) *The Age of Migration*. Palgrave MacMillan, Guilford.

26 Chappell, R, Teague, A and Diamond, I (1997) *Methodological research for a 'one number census'*. Office for National Statistics, Census Division. Available at: www.ons.gov.uk/census/2011-census/consultations/user-adv-groups/census-adv-groups/statistical-dev/meth-res-1-num-cen.pdf

27 Simpson, L (2007) 'Fixing the population: from census to population estimate'. *Environment and Planning A* (39) 1045-1057.

28 Abbott, O (2007) '2011 census coverage assessment and adjustment strategy'. *Population Trends* 127: 7-14.

29 ABS (2009) ABS 2006 Census Rebasing Papers: Estimating Statistical Local Area (SLA) and Indigenous Estimated Resident Population (ERP). *Internal paper; unpublished*.

30 Qvist, J (1999) 'Problems of coverage in the register of total population (RTB): Estimation of overcoverage by an indirect method'. *Statistics Sweden,* R&D Report 1999:1.

31 Brown, J, Abbott, O and Diamond, I (2006) 'Dependence in the 2001 one-number census project'. *Journal of the Royal Statistical Society*169: 883-902.

32 Dick, P and You, Y (1997a) 'Bayes and census undercoverage'. Statistical Society of Canada, Annual Conference; Proceeding of the Survey Methods Section, Fredericton, New Brunswick.

33 Dick, P and You, Y (1997b) 'A hierarchical Bayes analysis of census undercoverage'. Proceedings of the Statistics Canada Symposium 199; Work Session on Advances in Imputation, Canada: Ottawa.

34 Elliot, MR and Little, RJA (2000) 'A Bayesian approach to combining information from a Census, a coverage measurement survey and Demographic analysis'. *Journal of the American Statistical Association* 95 (450) 351-362.

35 Elliot, MR and Little, RJA (2005) 'A Bayesian approach to 2000 census evaluation using ACE survey data and demographic analysis'. *Journal of the American Statistical Association* 100 (470) 380-388.

Change in living arrangements following death of a partner in England and Wales, 1971 to 2001

Michael Hirst and Anne Corden

University of York

Abstract

Understanding trends and changes in the circumstances of couples separated by death is important for policy initiatives to reduce vulnerabilities associated with end of life care and for those who live on. This article uses widow(er)hood statistics and census data from the Office for National Statistics Longitudinal Study. It examines changes in couples' living arrangements and households at four successive censuses from 1971 to 2001and shows how these differ by age and gender on the death of a spouse or partner. Findings draw attention to the effects of ageing and mortality improvements as well as wider social and economic trends in family and household formation, residential independence in older age, and policy developments on long-term care provision for older people.

Contents

Introduction

Ageing and mortality are major challenges for policy and political debate[1]. Accurate projections are required to inform health and social care provision, pension reform, and insurance based products such as mortgage protection and equity release schemes. Trends in ageing and mortality are also changing the needs and circumstances of dying and bereaved people.

Death of a spouse or partner represents perhaps the most convergent set of circumstances that shape experiences of both mortality and bereavement. Key determinants of partnership formation such as age, income and education influence mortality rates and changes in mortality rates. Moreover, what happens to people when their spouse or partner dies is closely linked to their circumstances and behaviour as a couple, and the life they shared together.

Couples' living arrangements have implications for the availability of care and support before and after the death of a partner. Wider consequences for the health and well-being of the partner who lives on and any children of the relationship include the financial and economic adjustments that follow the death[2]. This article examines trends and changes in the family and household composition of couples separated by death, and provides descriptive analyses by age and sex of the surviving partner. We also provide provisional assessments of how many couples were living apart and how many unmarried couples were cohabiting before separation by death.

Methods

We used data from official mortality statistics and from the Office for National Statistics Longitudinal Study (ONS LS) at four successive census points from 1971 to 2001[3]. The LS links census and vital registration records for around one per cent of the population of England and Wales[4]. The original sample comprising approximately 500,000 individuals or 'LS members' was drawn from the 1971 census using four birth dates. Details of their circumstances have been updated with information collected in the 1981, 1991 and 2001 censuses. The sample is continually refreshed by the addition of one per cent of immigrants and births using the same birth dates, with further details of new entrants from subsequent censuses, and so remains nationally representative. Major strengths of the LS are its large sample size, low non-response and attrition rates, and inclusion of the institutional population. Limitations include the ten year gap between census enumerations, the restricted range of information collected at each census, changes over time in the definition and coding of key variables such as marital status, and differences between censuses in coverage and data quality.

Bereaved spouses are routinely identified by a trawl of the annual death files and linkage to the LS relies on matching their date of birth in the census record and the death register[5]. The register records legal marital status so death of a partner in the LS identifies widows and widowers, and does not cover cohabiting couples separated by death[6]. For the analysis described here, we constructed study groups before and after the death of a spouse, labelled pre-bereavement and post-bereavement cohorts respectively. Pre-bereavement cohorts included LS members present at a census whose spouse died within two years following that census; post-bereavement cohorts included LS members present at a census whose spouse had died less than two years before that census. A two year window either side of a census was a compromise between the need for sufficiently large cohorts for sub-group analysis, and the need to ensure that couples' circumstances reported in a census were not too far removed from those on separation by death.

Pre-bereavement and post-bereavement cohorts were constructed from LS samples covering the 1981, 1991 and 2001 censuses; only a pre-bereavement cohort could be identified in 1971 because prior widow(er)hoods had not been added to the initial sample. **Table 1** shows the number of people in each cohort. These comprised LS members usually resident in private households or communal establishments including, for example, a care home, hospice or hospital (apart from staff and relatives of staff)[7]. Sample numbers in the analysis varied because of missing or imputed information on relevant variables.

Table 1 Number of bereaved LS women and men in the study design

Census	Cohort	Women	Men	Total
1971	Pre-bereavement	3120	1446	4566
1981	Pre-bereavement	2773	1430	4203
	Post-bereavement	2595	1282	3877
1991	Pre-bereavement	2836	1347	4183
	Post-bereavement	2688	1229	3917
2001	Pre-bereavement	2644	1267	3911
	Post-bereavement	2631	1133	3764

Source: ONS Longitudinal Study

We evaluated the composition of bereavement cohorts for the impact of attrition and length of follow-up between the census and a spouse's death. We also examined apparent discrepancies in spouses' birth dates between the census record and the death register. Further details of the study design and these assessments are described elsewhere[8].

Findings

In this section, we use census data collected from 1971 to 2001 to examine trends and changes in the circumstances of married couples at each census who were separated by death within the following two years. These findings are based on the socio-demographic profiles of LS members in the pre-bereavement cohorts described above. Where appropriate, we draw comparisons between their circumstances and those of LS widows and widowers at each census, from 1981 on, whose spouses had died during the previous two years. Comparisons between pre-bereavement and post-bereavement cohorts provide provisional indications of changes that might be associated with, or accompany, the death of a spouse. We then go on to evaluate the coverage of LS widow(er)hood statistics, which are based on death of a spouse, and estimate the number of couples, both married and cohabiting, separated by death in 2001.

Age and sex

Death of a spouse is predominantly experienced by older people and now happens much later in the life course than in previous generations. In the early years of the 21st Century, death of a spouse typically occurred when people were in their early 70s, an increase of around six or seven years on three decades earlier (**Table 2**). The ageing of spousal bereavement may be due to increased health benefits associated with marriage as well as mortality improvements affecting all groups[9]. The additional years lived by married people shown here exceed estimated increases of two to three years since 1981 in life expectancy for the general population at age 65[10].

Table 2 Mean and median age of LS women and men on death of spouse, 1971–2001, England and Wales

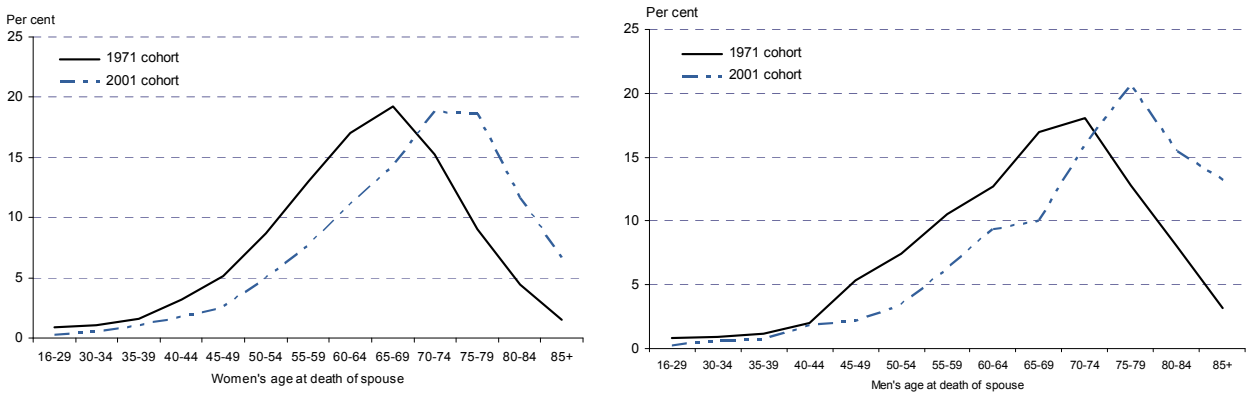
	1971	1981	1991	2001	1971 – 2001 Change (years)	1971 – 2001 Change (per cent)
Women						
Mean age	63.1	65.5	68.1	69.3	6.2	9.8
(SD)	(11.5)	(11.5)	(11.3)	(11.7)		
Median age	64	67	69	71	7.0	10.9
Men						
Mean age	65.6	67.8	70.1	71.9	6.3	9.6
(SD)	(12.1)	(12.3)	(11.8)	(12.4)		
Median age	67	70	71	74	7.0	10.4
Base numbers						
Women	*3120*	*2773*	*2836*	*2644*	–	–
Men	*1446*	*1430*	*1347*	*1267*	–	–

Source: ONS Longitudinal Study

Sex differences in spousal bereavement reflect the tendency for women to marry men older than themselves. Women were typically three years younger than their husband although marital age differences of two years or less were most common across the bereavement cohorts. Similarly, women were bereaved around two to three years younger on average than their male counterparts (Table 2). For women and men alike however, increases in the age at which couples were separated by death have boosted the proportion of people bereaved in their 70s and beyond (**Figure 1**). In the 24 months following the 2001 census, 57 per cent of women were aged 70 and over when their spouse died, up from 33 per cent in an equivalent period following the 1971 census. Comparable figures for bereaved men were 64 and 44 per cent respectively.

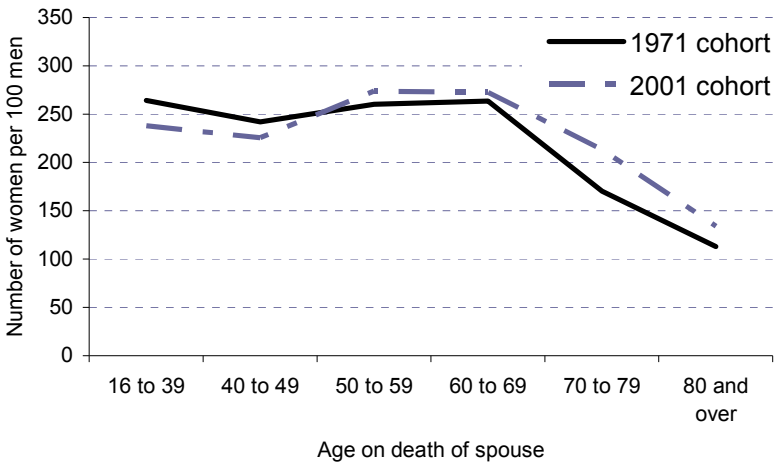
The age difference at marriage and women's longer life expectancy largely account for the predominance of widows among bereaved spouses. Overall, twice as many women as men experienced the death of a spouse and while that ratio fluctuated a little, there was no firm trend across these cohorts. Sex differences in spousal bereavement were substantial in all but the oldest old, such that men in their 80s and older were almost as likely as women to experience the death of a spouse (**Figure 2**). However, the ratio of widows to every 100 widowers appears to have increased in older age groups since the early 70s although this may reflect improvements in the identification of LS women following death of a spouse[5].

Figure 1 **Percentage age distribution of LS men and women on death of spouse, 1971 and 2001, England and Wales**



Source: ONS Longitudinal Study

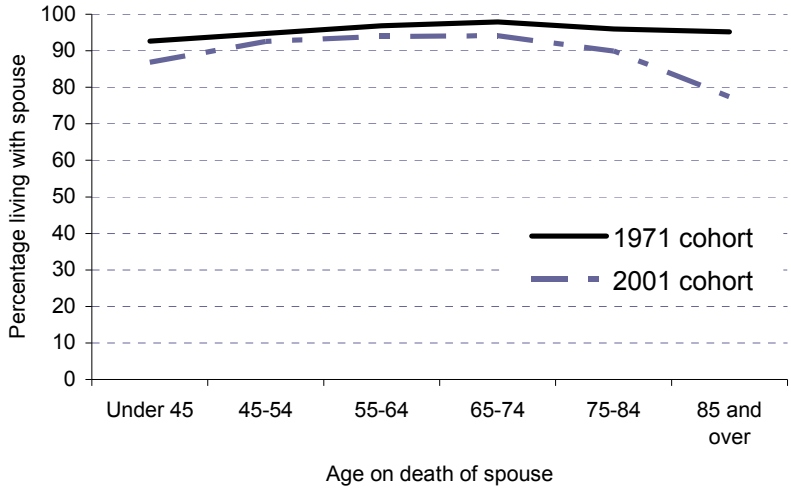
Figure 2 **Sex ratio of LS members by age on death of spouse, 1971 and 2001, England and Wales**



Living arrangements

As might be expected, most people were living with their marital partner at the census before their spouse died. That proportion declined between 1971 and 2001, from 96 to 91 per cent, especially in older age groups (**Figure 3**). Several factors would account for couples apparently not living together before one spouse died. Their older age profile might indicate that admission of the LS member to institutional care was one such reason (see further below). The contribution of other factors is less clear as the circumstances of 'absent spouses' are not recorded. Some couples may have lived apart because they had separated though still formally married. Some spouses may have entered institutional care; they would not have been counted as co-resident even if they lived in the same establishment as the LS member to whom they were married. Some cohabiting couples and LS members apparently living alone may have married their partner shortly after the census, before separation by death. However, spouses working away from home or serving in the armed forces should have been treated as usually co-resident.

Figure 3 LS members by age living with a spouse before separation by death, 1971 and 2001, England and Wales



Source: ONS Longitudinal Study

Information on living arrangements gathered in 1991 and 2001 shows very few couples living as unmarried cohabitants before the death of a spouse; there were probably fewer in earlier decades when cohabitation was less widespread (**Table 3**). Apart from married couples then, people were mostly living singly at the census before the death of a spouse and their numbers increased slightly between 1991 and 2001. Their partnership status is uncertain although all had, or would have, a spouse who died within two years of the census.

Table 3 shows further that few people in the post-bereavement cohorts had apparently re-partnered in the first year or so after their spouse died[11]. Bereaved men were somewhat more likely to be married or cohabiting than bereaved women although women and men under 65 were more likely to have re-partnered than those who were older (women 3.1 and 0.8 per cent respectively, men 5.6 and 1.9 per cent respectively in the 2001 cohort). Re-partnering after the death decreased between the 1991 and 2001 cohorts, associated with a decline in formal re-marriage. However, partnership rates remained steady among bereaved women under 65 (3.2 and 3.1 per cent in 1991 and 2001 respectively) but declined among men (8.4 and 5.6 per cent respectively). Among bereaved people aged 65 and over, partnership rates fell by half between the two cohorts for women and men alike.

As noted above, some couples were living apart before separation by death because one or both spouses had entered institutional care. The extent of institutional living, for LS members only, is shown in **Table 4**, which mirrors wider trends pointing to a rise in admissions during the 1980s followed by a levelling off thereafter. Death of a spouse may itself influence the volume, pattern and timing of institutional admissions. Although few people were recorded in such settings, more LS members lived in communal establishments after their spouse died than in the pre-bereavement cohorts.

Table 3 Living arrangements of LS women and men before and after death of spouse, 1991 and 2001, England and Wales

Percentages	Pre-bereavement		Post-bereavement	
	1991	2001	1991	2001
Women				
Single/lone parent	4.3	5.9	97.5	98.5
Married couple	95.2	93.3	2.1	0.9
Cohabiting couple	0.5	0.7	0.3	0.6
Men				
Single/lone parent	3.8	5.7	94.8	97.2
Married couple	95.6	92.5	3.7	1.7
Cohabiting couple	0.7	1.8	1.4	1.1
Base numbers				
Women	*2788*	*2599*	*2610*	*2531*
Men	*1327*	*1248*	*1178*	*1087*

Note: percentage LS members in private households; data on cohabitation is not available for 1971 and 1981.

Source: ONS Longitudinal Study

Table 4 LS women and men living in communal establishments before and after death of spouse, 1971–2001, England and Wales

Percentages	1971	1981	1991	2001
Women				
Pre-bereavement	0.3	0.4	1.5	1.7
Post-bereavement	–	1.8	2.7	3.7
Men				
Pre-bereavement	0.2	0.8	1.3	1.4
Post-bereavement	–	1.4	4.1	4.1
Base numbers				
Women				
Pre-bereavement	*3119*	*2773*	*2836*	*2643*
Post-bereavement	*n/a*	*2594*	*2688*	*2629*
Men				
Pre-bereavement	*1446*	*1429*	*1347*	*1266*
Post-bereavement	*n/a*	*1280*	*1229*	*1133*

Source: ONS Longitudinal Study

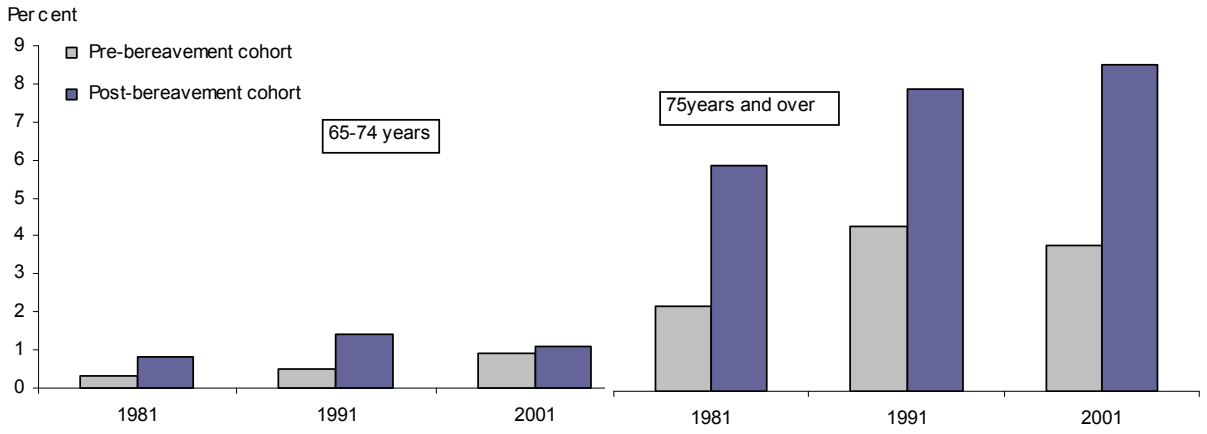
Almost all the LS members living in communal establishments (99 per cent), in both pre-bereavement and post-bereavement cohorts, were aged 65 and over; however, the oldest old were most likely to be recorded in such settings. People aged 75 and over also showed the largest uptake, relative to pre-bereavement levels, in institutional living after the death of a spouse (**Fig 4**).

Household composition

Trends in couples' living arrangements before separation by death have contributed to a decline in multi-generational households. Since the 1980s such couples were less likely to be looking after children and young people under 16, or young people under 18 and still at school or college. Trends in the proportion of couples living with parents were less clear-cut but point to a decline in

the 1990s (**Table 5**). The increasing age of spousal bereavement may have influenced both these trends, meaning that any children were more likely to have been counted as adults and left the parental home, and parents of older couples were more likely to have died or entered institutional care.

Figure 4 LS members by age living in communal establishments before and after death of spouse, 1981–2001, England and Wales



Source: ONS Longitudinal Study

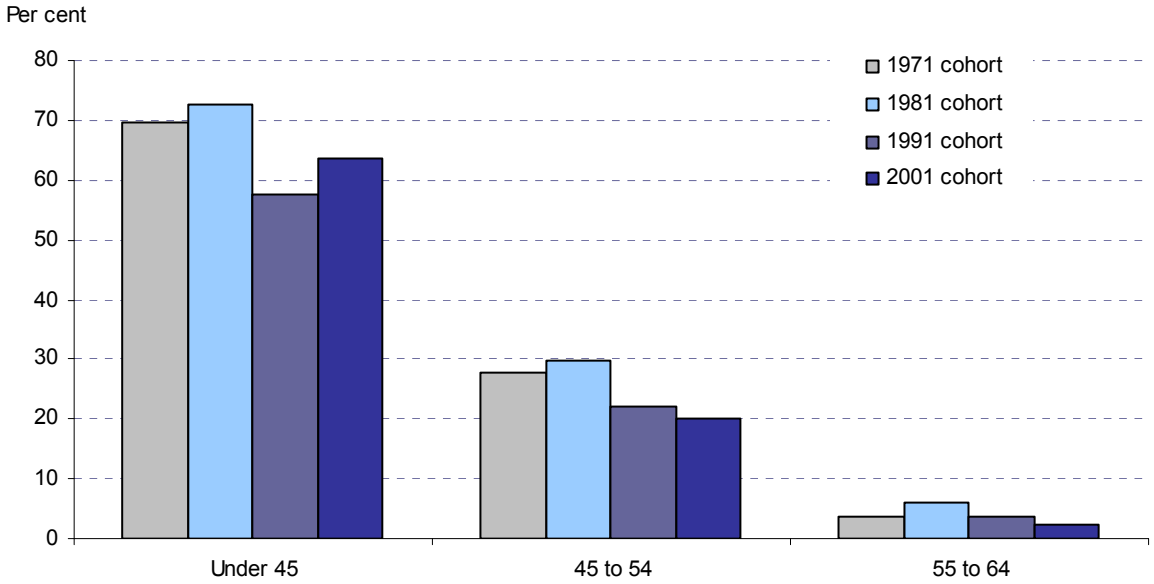
Table 5 LS women and men living with dependent children or parents before death of spouse, 1971–2001, England and Wales

Percentages	**1971**	**1981**	**1991**	**2001**
Dependent children in the family				
Women	18.5	18.3	14.4	13.1
Men	22.5	26.8	17.2	17.9
Parents in the household				
Women	n/a	2.8	2.4	0.9
Men	n/a	3.1	2.9	0.0
Base numbers				
Women	*1667*	*1266*	*1023*	*780*
Men	*640*	*523*	*408*	*302*

Note: percentage LS members under 65 in private households; includes step children and step parents in 2001 only.
Source: ONS Longitudinal Study

Despite an overall trend towards single generation households, family composition when one spouse died continued to reflect typical life cycle stages. Most people under 45, and sizeable minorities of those between 45 and 54, would have had responsibility for looking after children and young people before and after their spouse died (**Figure 5**). However, the proportion of younger couples with children fluctuated somewhat between 1971 and 2001 and showed a net decline within the age groups used here.

Figure 5 LS members by age living with dependent children before death of spouse, 1971–2001, England and Wales



Note: percentage LS members in private households.

Source: ONS Longitudinal Study

At the same time, there was a marked decline in people living with adult children before their spouse died: the proportion in such households in 2001 was almost half that of the early 1980s (**Table 6**). The impact was most noticeable among the oldest old, reflecting the increasing trend for older people not to live in the same household as their adult children (**Figure 6**). As might be expected, couples aged 45 to 54 were most likely to be living with adult children indicating that death of a spouse in middle age often occurred when children were entering adult life and relationships, and establishing their own working careers.

Death of a spouse may itself influence inter-generational living, perhaps linked to the support and accommodation needs of the one who lives on. Findings show an increased propensity, across the age range, to live with adult children after the death. How that situation arose – who moved and when – cannot be determined from these data. Table 6 shows that differences in proportions between pre-bereavement and post-bereavement cohorts have changed little since the early 1980s, indicating that the impact of a spouse's death on sharing accommodation with adult children has not diminished. With each successive cohort however, fewer people were living with adult children following the death of a spouse.

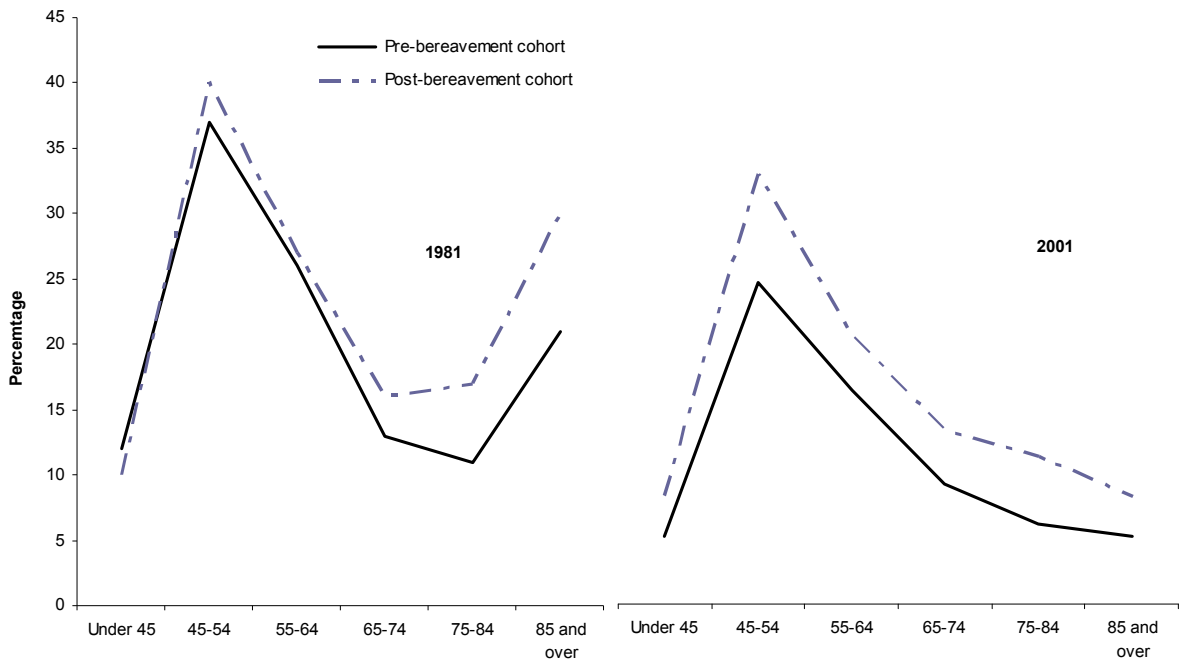
Table 6 LS women and men living with adult children before and after death of spouse, 1981–2001, England and Wales

	1981	1991	2001
Women			
Pre-bereavement	19.4	16.6	10.8
Post-bereavement	22.3	19.9	14.6
Men			
Pre-bereavement	17.3	16.2	9.1
Post-bereavement	19.2	17.8	13.5
Base			
Women			
Pre-bereavement	*2761*	*2793*	*2354*
Post-bereavement	*2548*	*2615*	*2459*
Men			
Pre-bereavement	*1418*	*1330*	*1140*
Post-bereavement	*1262*	*1179*	*1053*

Note: percentage LS members in private households; data not available for 1971.

Source: ONS Longitudinal Study

Figure 6 LS members living with adult children before and after death of spouse by age, 1981 and 2001, England and Wales



Note: percentage LS members in private households.

Source: ONS Longitudinal Study

Household size

Changes in living arrangements and household formation have contributed to a rise in the extent to which couples lived on their own before separation by death. The proportion of two person households rose by over ten percentage points in the thirty years to 2001, with much of the increase occurring in the 1970s and 1980s (**Table 7**). There was a much smaller increase in the number of people living on their own probably reflecting, as suggested above, admission of a spouse into institutional care.

As a consequence, a substantial majority of people lived alone following the death of a spouse. Living alone was most prevalent among older widows and widowers and has become more widespread with each succeeding cohort (**Figure 7**).

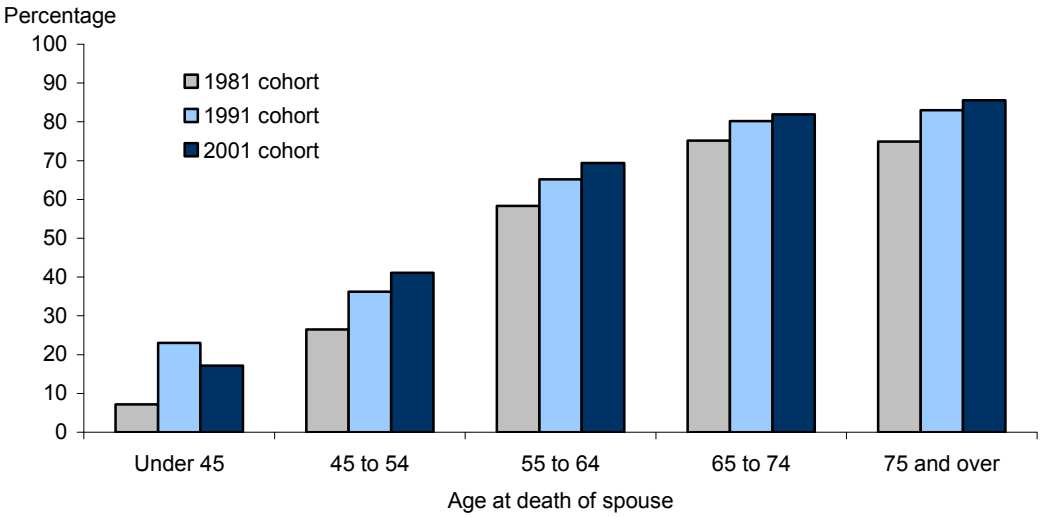
Table 7 LS women and men by number of people in household before and after death of spouse, 1971–2001, England and Wales

	Pre-bereavement				Post-bereavement		
	1971	**1981**	**1991**	**2001**	**1981**	**1991**	**2001**
Women							
1	0.7	0.9	2.9	4.2	63.0	72.4	75.7
2	64.0	69.6	74.7	75.5	21.3	17.1	14.8
3	19.4	17.0	14.1	13.0	8.5	6.3	6.0
4 or more	15.9	12.4	8.3	7.2	7.3	4.3	3.5
Men							
1	0.8	1.6	2.8	4.7	63.2	71.6	76.2
2	63.1	70.4	74.5	76.0	20.0	17.0	14.6
3	20.4	15.4	14.4	11.5	9.0	5.6	5.4
4 or more	15.8	12.6	8.3	7.8	7.8	5.8	3.8
Base							
Women	*3111*	*2761*	*2793*	*2599*	*2548*	*2615*	*2531*
Men	*1443*	*1418*	*1330*	*1248*	*1262*	*1179*	*1087*

Note: percentage LS members in private households; post-bereavement cohort not available in 1971.

Source: ONS Longitudinal Study

Figure 7 Percentage of LS members by age living alone after death of spouse, 1981–2001, England and Wales



Note: percentage LS members in private households.

Source: ONS Longitudinal Study

Residential mobility

Changes in living arrangements and household composition following the death of a spouse may result from people moving to join another household or being joined by new household members. Moving house may be linked to changes in people's accommodation and support needs: to downsize, release equity, or to be nearer friends and family, for example. Findings indicate that such moves were not widespread and may have decreased over time. However, it was not possible to count recent moves into institutional care which can be identified in the LS only by comparing residential location at successive censuses.

Table 8 shows that bereaved people were somewhat more likely to have moved house in the past year than those living in private households before separation by death. These comparisons may overstate the impact of a spouse's death on moving house because people in post-bereavement cohorts could have moved before the death if their spouse had died less than 12 months before a census, that is, within the set period for reporting a change of address. Taking account of this time frame did not alter the broad conclusion that residential mobility was probably more widespread after the death. People whose spouse died more than 12 months before a census were somewhat more likely to have moved house in the past year than those whose spouse died within 12 months of a census (6.2 and 4.9 per cent respectively in the 2001 post-bereavement cohort).

Table 8 LS women and men who changed address in year before and after death of spouse, 1981–2001, England and Wales

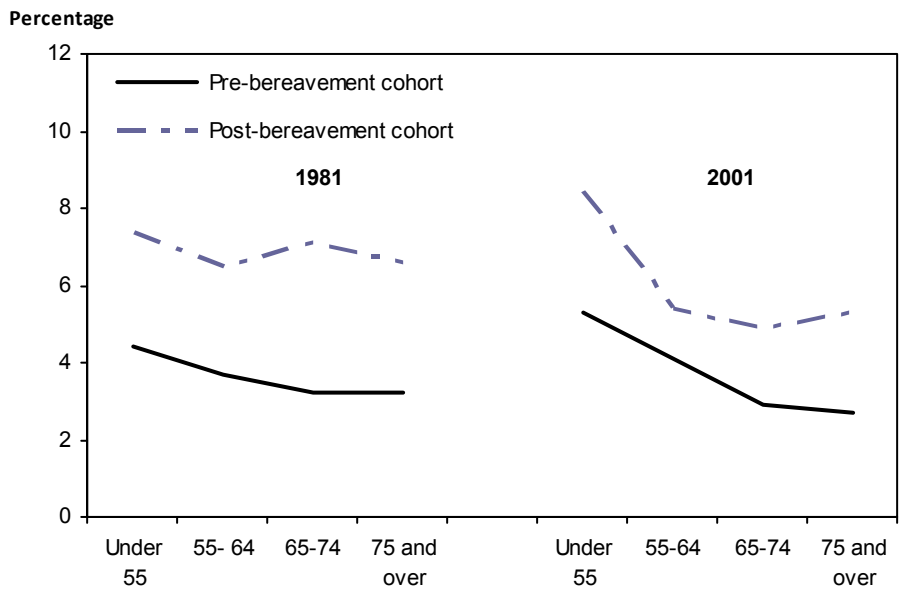
	1981	1991	2001
Women			
Pre-bereavement	3.5	3.2	3.3
Post-bereavement	7.1	4.6	5.8
Men			
Pre-bereavement	3.5	1.7	3.5
Post-bereavement	6.3	4.7	5.3
Base numbers			
Women			
Pre-bereavement	*2761*	*2793*	*2421*
Post-bereavement	*2548*	*2615*	*2374*
Men			
Pre-bereavement	*1418*	*1330*	*1204*
Post-bereavement	*1262*	*1179*	*1025*

Note: percentage LS members in private households; data not available for 1971.

Source: ONS Longitudinal Study

Trends in residential mobility may be influenced by changes in the housing market, including legislation in the 1980s which enabled public sector tenants to purchase their home at discounted prices. From 1981 to 2001, residential mobility increased somewhat among people under 55 and declined slightly in older age groups (**Figure 8**). These findings may reflect increased home ownership in older age groups and preferences for staying put, as well as the often drawn out process of buying and selling a house. Home owners were half as likely as those in rented accommodation to have moved house following the death of a spouse prior to the 2001 census: an odds ratio of 0.48 compared with 0.86 and 0.57 in the 1981 in 1991 post-bereavement cohorts respectively.

Figure 8 LS members by age who changed address in the year before and after death of spouse, 1981–2001, England and Wales



These findings shed some light on the timing of residential moves following the death of a spouse. One year migration rates in the 1991 and 2001 cohorts were around five per cent, only marginally above pre-bereavement levels (Table 8). Though not directly comparable, more than a third of widows and widowers aged 50 and over changed their address between 1991 and 2001[12]. These comparisons between one year and ten year migration rates indicate that most residential moves following the death of a spouse probably occurred several years after the death, rather than in the first 12 months or so.

Estimating the number of couples separated by death

So far we have focused on married couples and the circumstances of LS members before and after their spouse died. Their identification was based on reported marital status when a death is registered. These data provide a useful starting point for estimating the incidence of spousal bereavement. However, they misrepresent the number of couples separated by death because the classification of marital status has yet to reflect the diversity of family forms. Separated, unmarried cohabitation, same sex partnerships, and 'living apart together' are not currently recognised in the registration process. As a consequence, some people recorded as married at the time of death may not have been in a relationship or not married to the partner with whom they were living; and some people recorded as single, widowed or divorced may have been living with a partner.

By examining the living arrangements of LS members before they died (rather than, as above, before their spouse died), it is possible to gain some indication of how many were married or living as married from census records. **Table 9** compares the census defined marital status of LS members who died within two years following the 2001 census with the marital status record on the registration of their death. The table is organised according to whether or not LS members said they were living with a partner at the time of the census, distinguishing between married and cohabiting couples.

Reports of marital status and partnership histories are not always reliable or easy to interpret and comparison of individuals' reported marital status in census and death records reveals both expected and unexpected shifts. Thus, Table 9 shows that 174 people in a married couple at the time of the census were recorded as widowed when they died indicating, quite plausibly, that their spouse may have died before them. Among cohabiting couples, a decrease in the number of divorced or single people, matched by an increase in the number of married people, similarly draws attention to a plausible change in their legal marital status before separation by death. Changes in marital status of those in communal establishments are perhaps least reliable because details registered on death may have been informed by members of staff rather than close relatives. They indicate a decrease in the number of married people that, on the face of it, was associated with an increase in those regarded as divorced or single, but no increase in the widowed group.

Differences in reported marital status between the census record and registration of death can be used, with caution, to inform assumptions about the marital status of couples separated by death. One set of assumptions (described in **Box one**) indicates that 4,391 people might have been part of a couple, though not necessarily co-resident, when they died. This figure is somewhat higher than the 4,301 recorded as married on death, suggesting that official mortality statistics represent

around 98 per cent of couples where one partner died. Other assumptions would produce different estimates.

That 98 per cent of couples separated by death were probably legally married may be compared with an estimated 96 per cent derived from the British Household Panel Survey (BHPS), explained in **Box two**. The true estimate may lie between these two figures. The LS figure is probably an overestimate because it excludes same sex couples. The BHPS is probably an underestimate because some cohabitants could have married before their partner died some six months on average after the interview. In addition, the BHPS does not identify couples where one or both partners were in a communal establishment at the date of death. Neither source fully counts couples that usually live apart.

Whatever the limitations of the LS and BHPS data sets, the number of couples where one partner dies is probably higher than indicated by those recorded as married when death is registered, mainly because registration fails to take account of unmarried cohabitation. In 2001 over 203,000 people in England and Wales were recorded as married on registration of their death[13]. Applying a midpoint estimate between the LS and BHPS estimates, that 97 per cent of couples separated by death were legally married, produces an additional 6,000 couples who were cohabiting when one partner died, making almost 210,000 couples altogether. These estimates provide a starting point for further investigation and comparison with other longitudinal data sets.

Box one Partnerships on separation by death

From Table 9, we may assume that those in a couple comprised the following:

- 3,890 people in a married couple at the census and recorded as married at death.
- 238 people in a cohabiting couple at the census. This may be an overestimate if cohabitation happens to be a relatively transitory arrangement.
- 54 people representing the net increase in the number of individuals and lone parents recorded as married between the census and date of death. This figure may underestimate the extent of re-partnering in this group through unmarried cohabitation.
- 209 married people who were living in communal establishments at the census.

These figures together indicate that 4,391 LS members might have been part of a couple when they died.

Table 9 Living arrangements of LS members in 2001 by marital status at census and at date of death, England and Wales

	April 2001 at census	Died before 1 May 2003
Married couple		
Single	0	5
Married	4089	3890
Widowed	0	174
Divorced	0	9
Separated	3	n/a
Not stated	n/a	14
Total	4092	4092
Cohabiting couple		
Single	51	46
Married	19	51
Widowed	50	51
Divorced	104	87
Separated	14	n/a
Not stated	n/a	3
Total	238	238
Individual or lone parent		
Single	743	739
Married	97	151
Widowed	3272	3268
Divorced	455	478
Separated	100	n/a
Not stated	n/a	31
Total	4667	4667
Communal establishment		
Single	288	307
Married	240	209
Widowed	1394	1375
Divorced	66	90
Separated	7	n/a
Not stated	n/a	14
Total	1995	1995
All living arrangements		
Single	1082	1097
Married	4445	4301
Widowed	4716	4868
Divorced	625	664
Separated	124	n/a
Not stated	n/a	62
Total	**10992**	**10992**

Notes:

'n/a' means the category was not available.

'Married' includes those who have remarried.

'Single' means never married.

Separated but legally married' could be recorded only in the census.

'Not stated' could be recorded only at registration of death.

Source: ONS Longitudinal Study

Box two Estimating the number of couples separated by death using the British Household Panel Survey

The BHPS is a nationally representative survey of private households and aims to interview the same people every year[1]. Panel deaths between successive interview waves from 1991 to 2004 identified 756 people whose partner died[2]. Of these, 27 people (four per cent) said they were cohabiting when interviewed before their partner's death, including two people who lived with partners of the same sex. These interviews took place around six months on average before the death (SD=4).

1. Lynn P (ed.) (2006) *Quality Profile: British Household Panel Survey Version 2.0: Waves 1 to 13: 1991-2003*. Institute for Social and Economic Research, University of Essex, Colchester.

2. Corden A, Hirst M and Nice K (2008) *The Financial Implications of Death of a Partner*. Social Policy Research Unit, University of York, York.

Conclusion

People whose spouse died in the early years of the 21st century were generally older than their counterparts three decades earlier. Over the same period, the number of married couples separated by death in England and Wales has progressively fallen[14]. Longer life expectancy has been a key driver of these trends although the closing gender gap in mortality improvements has had little impact so far on reducing the predominance of widows among bereaved spouses. For men and women alike however, death of a spouse is mostly experienced in the context of a long life course and increasingly shaped by the circumstances and transitions associated with older age.

The ageing of spousal bereavement has, in turn, influenced couples' family and household circumstances before separation by death. Fewer couples were living with adult children, younger children or elderly parents. Across the study period, couples were increasingly living on their own before separation by death, accompanied by a small but rising number of older people living alone or in communal establishments before their spouse died. Among other factors, wider trends in family and household formation, residential independence in older age, and policy developments on long-term care provision for older people have shaped changes in couples' circumstances. Moreover, stage in the family life cycle was an important determinant of couples' living arrangements before one spouse died although examination of partnership histories was beyond the scope of the data used here.

Following bereavement, most people lived alone, especially in older age groups, signifying for many the adjustment to loss and changes to status or role that follow the death of a spouse. For some, entry to institutional care or living with adult children may be triggered by the death of a spouse. Although residential mobility appears to be associated with widow(er)hood, moving house may not take place until several years after the death, prompted by less direct or immediate consequences of bereavement including changes in health and mobility, income or accommodation needs.

The number of couples separated by death is expected to rise in coming decades as the post-war baby boomer generations enter older age groups. Population projections show deaths increasing throughout much of the 21st Century[15], while the number of people aged 65 and over living in married or cohabiting couples is expected to increase from under five million to over eight million between 2007 and 2031[16].

Needs for practical and emotional support for dying and bereaved people may increase accordingly, fuelling debate about risk factors for complicated or prolonged grief and targeting effective interventions. Trends in ageing, mortality and living arrangements may mean more couples reaching older age with diminished resources and capabilities to manage the care and support needs of a partner approaching the end of life, and cope with the transitions and adjustments that follow. Death, dying and bereavement present clear entry points for limiting or preventing social, economic and health vulnerabilities[17]. Additionally, a new legal framework would help protect the property rights and financial well-being of unmarried and unregistered couples, their children and other home sharers[18].

Trends and changes in couples' living arrangements before separation by death are key elements in projecting demand for and spending on adult social care, welfare assistance, and long-term care for older people. Information on their circumstances is important for policy development, service delivery and professional practice to support people through this period.

Key findings

Based on married couples at each census, 1971 to 2001, who were separated by death within the following two years, and the circumstances of widows or widowers at each census, from 1981 on, who had been bereaved during the previous two years, this study found that:

- Age on death of a spouse increased by almost seven years between the early 1970s and the first years of the 21st Century (from around 65 to 72 median ages). Nearly three out of four people (72 per cent) were aged 65 and over when their spouse died, up from 52 per cent three decades earlier.
- Couples increasingly lived on their own before separation by death: 76 per cent of couples in private residences reported no other household members in 2001 compared with 64 per cent in 1971.
- Most widow(er)s lived alone after their spouse died and were increasingly likely to do so (76 per cent in 2001, up from 63 per cent in 1981).
- Around two out of three people under 45 were looking after dependent children at the census before their spouse died although their numbers declined in absolute terms between 1971 and 2001.
- The proportion of couples not sharing the same address at the census before one spouse died increased from three to nine per cent between 1971 and 2001, including couples where one or both members lived in communal establishments, as well as couples who had separated or were soon to be married.

Key findings

- Most residential moves following the death of a spouse probably took place several years later rather than in the first 12 months or so after the death.
- An estimated three per cent of couples were not formally married to each other when separated by death.

Acknowledgements

The permission of the Office for National Statistics to use the Longitudinal Study is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information and User Support (CeLSIUS). The authors wish to thank especially: Julian Buxton, Research and Computing Support Officer at CeLSIUS, for assistance in preparing data sets and offering advice on the analysis; and Rachel Stuchbury, Research and Computing Support Officer at CeLSIUS, for reading an earlier draft of this article and offering valuable comments. CeLSIUS is supported by the UK Economic and Social Research Council Census of Population Programme (Award Ref: RES-348-25-0004). Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. This study was approved by the ONS Longitudinal Study Research Board (Project 30093B). The authors alone are responsible for the analysis and interpretation of the data reported here.

References

1 Dunnell, K (2008) ‘Ageing and mortality in the UK: National Statistician’s annual article on the population’. *Population Trends* 134: 6-23.

2 Corden A, Hirst M, Nice, K (2010) ‘Death of a partner: financial implications and experience of loss’. *Bereavement Care* 29: 23-28.

3 Brassett-Grundy, A (2003) *LS User Guide 20: Researching Households and Families Using the ONS Longitudinal Study.* Office for National Statistics and Centre for Longitudinal Studies, Institute of Education, University of London, London.

4 Blackwell, L, Lynch, K, Smith, J, Goldblatt, P (2003) *Longitudinal Study 1971–2001: Completeness of Census Linkage*. Series LS 10, Office for National Statistics and Centre for Longitudinal Studies, Institute of Education, University of London, London.

5 Widow(er)hoods by sex 1971 to 2004. Available at: www.ons.gov.uk/about/who-we-are/our-services/longitudinal-study/data-quality/event-sampling-and-linkage/widow-er--hoods/index.html#Equation

6 Deaths of registered same sex partners have been recorded in the LS following implementation of the Civil Partnership Act (2004), which legally recognised registered same sex partnerships in the UK.

7 Cases where there was a discrepancy between the main LS and widow(er)hood files in the sex of the deceased spouse were excluded. The 2001 bereavement cohorts also exclude cases with imputed age values, students not at their term-time address, and LS members with initial extract records only.

8 Hirst, M, Corden, A (2010) *Trends and Changes in Household and Personal Circumstances on Death of a Partner in England and Wales, 1971 to 2001*. Social Policy Research Unit, University of York, York.

9 Murphy, M, Grundy, E, Kalogirou, S (2007) 'The increase in marital status differences in mortality up to the oldest age in seven European countries, 1990-99'. *Population Studies 61*: 287-298.

10 Office for National Statistics (2006) 'Health expectancies in the UK, 2002'. *Health Statistics Quarterly* 29: 59-62.

11 Some married and cohabiting couples in the post-bereavement cohorts could have been living together before the LS member's spouse died.

12 Uren, Z, Goldring, S (2007) 'Migration trends at older ages in England and Wales'. *Population Trends* 130: 31-40.

13 Office for National Statistics (2003) *Deaths by age, sex and marital status, 2001*. Table 9 in Series DH1, no 34.

14 New widow(er)s statistics for England and Wales, 1939 to 2003, are compiled by ONS from mortality statistics and published as follows: New widowers data set (PMH61A) available at: www.statistics.gov.uk/StatBase/xsdataset.asp?vlnk=5287&Pos=&ColRank=1&Rank=224 New widows data set (PMH61B) available at: www.statistics.gov.uk/StatBase/xsdataset.asp?vlnk=5288&Pos=1&ColRank=1&Rank=208

15 Office for National Statistics (2009) 2008-based National Population Projections. Office for National Statistics, London. Available at: www.statistics.gov.uk/StatBase/Product.asp?vlnk=8519

16 Office for National Statistics (2009) '2006-based marital status and cohabitation projections for England and Wales'. *Population Trends* 136: 112-120.

17 Department of Health (2008) *End of Life Care Strategy. Promoting High Quality Care for All Adults at the End of Life*. Department of Health, London.

18 Law Commission (2007) Cohabitation: the Financial Consequences of Relationship Breakdown (Cm. 7182). The Law Commission, London.

An analysis of patient register data in the Longitudinal Study – what does it tell us about the quality of the data?

Steve Smallwood and Kevin Lynch

Office for National Statistics

Abstract

This article uses data for members of the ONS Longitudinal Study (LS) from both Census 2001 enumeration and patient registrations "frozen" on census day 2001 from the National Health Service Central Register (NHSCR) to examine potential sources of difference in area of usual residence.

Overall 95.7 per cent of ONS LS members enumerated at census resided in the same area as recorded on the NHSCR data. Where areas differed, or the ONS LS member was not on the NHSCR on census day, subsequent NHSCR records were examined. Records flagged on the NHSCR as ONS LS members in England and Wales on census day but with no census record were also investigated.

Contents

Introduction

Patient register information is a key administrative source used in the construction of population estimates[1]. It is also used in the construction of small area population figures[2]. The main use of this information is for estimating internal migration between local authorities. Previous work has considered aggregate (macro) level comparisons of patient register and census data[3]. However a systematic analysis of linked (micro level) data analysis has not been carried out. The micro comparisons here help in understanding the quality of the information used and may provide lessons for rebasing the mid-year estimates after the 2011 Census. Increasingly, however, administrative sources are being considered for use in constructing demographic information. The Department of Health are also considering making increased use of the information they hold for resource allocation; they therefore have an interest in the quality of the patient register information. This article uses the one data source that links information from patient registration to census data, the ONS Longitudinal Study.

The ONS Longitudinal Study

The ONS Longitudinal Study (LS) contains linked census and vital event data for one per cent of the population, selected using dates of birth, for England and Wales. Information from the 1971, 1981, 1991 and 2001 Censuses has been linked together, along with information on events such as births, deaths and cancer registrations. At each census, data on more than 500,000 sample members are included. During the 30 years of the study, over one million people have been recorded in the sample at some point.[4]

The ONS LS was set up in the 1970s to meet the need for better data on mortality and fertility. Since then it has been used to address a wide range of research questions including studies of social mobility, ageing and migration. Information on migration comes from patient registration data on the National Health Service Central Register (NHSCR)[5] accessed using the Central Health Register Inquiry System (CHRIS). The paper uses some terms which relate to the data on the NHSCR which may not be familiar to readers; these are described in **Box one**.

The approach

The comparison of 2001 Census data with information on patient registration from the NHSCR provides a unique opportunity to see where the census placed people in the population compared with the patient register data. The ONS LS now has NHSCR data up to 2007 and so it is possible to get a good picture of how the census compares with the GP register data and subsequent changes through the decade. There are, however, some limitations to the analysis in that moves on the NHSCR are only picked up between Health Authorities (HAs), the former health areas before the introduction of Primary Care Organisations in 2001[6]. It should be noted that this comparison is between the health area where people were actually resident, based on census enumeration address, with the area in which they were registered with a GP. For sub-national migrants the population estimates will contain similar time lags and ‘errors’ as the NHSCR and other NHS systems giving GP registration data, as sub-national migration moves are derived from those data sets.

Box one
Terminology used for National Health Service Central Register data

Posting

This term refers to the move of a patient's registration from one health authority area to another.

Cancellation

This refers to the removal of a patient from the NHSCR – this will normally be because a General Practitioner has notified that the person has left their practice and can no longer be traced.

Embarkation

This refers to patient records where the patient has notified the doctor that they are moving abroad. Technically anyone who goes abroad for more than three months should notify their doctor.

Moves between health authority areas clearly do not cover all moves, but they do cover around four in five moves between local authorities. Examining the moves in the ONS LS Census data in the year prior to 2001 Census (using the address one year ago question) 78.5 per cent of moves across local authority boundaries were also moves between health authority areas. This compares with a very similar figure of 79.4 per cent for moves in the period mid-2000 to mid-2001 from the data used to create internal migration moves for mid-year estimates. Note that there will be a small number of cases in the analysis in this paper where NHSCR picks up where the patient registers with a doctor in a different health authority area but the move was within a local authority area.

Information on five different groups of people was examined. These groups were:

(1) Those present at 2001 Census[7] who also have an NHSCR record for the same area current at the 2001 Census date

For this group we can assume that the census, and therefore mid-year population estimates (MYEs) based on the census, have the person in the correct place. Although place here refers to health authority area, it does not necessarily mean people are recorded at the same address as the LS does not hold NHSCR address data. Note that it is also possible for individuals to be at the same home address on both the census and the GP register, but the GP surgery may be across a health area boundary.

(2) Those present at 2001 Census but whose NHSCR record at census day in 2001 is in an area other than their Census address

For this group it seems reasonable to assume that the census has the person in the correct place but that the GP register does not, as Census is completed at a particular point in time. Of course, if these people subsequently move internally on the NHSCR then the MYEs may wrongly move

those people into an area where they have already been placed by the census (taking them out of an area in which they were not enumerated).

(3) Those present at 2001 Census but whose NHSCR record had been cancelled prior to census day and no active record exists at 2001

For this group it can be assumed that the person is missing from the NHSCR, either because they have been 'wrongly' removed, correctly removed but failed to re-register, or that they are legitimately not on the register as they are covered by other medical arrangements, for example, that they are in the armed forces.

(4) Those present at 2001 Census but not appearing until later on the NHSCR, if ever

This group is likely to consist of migrants who have not registered with the NHS at the time of the census and who may or may not register later.

(5) Those absent at 2001 Census for whom an NHSCR record existed at census time

Many of this group will be census 'non response' or will be LS members who have replied to census but the information could not be linked; however, it will also include those who have left England and Wales[8] but not notified their GP that they have left.

In order to get an idea of the importance of each of these groups they need to be seen in the context of the total population. For categories 1 to 4 above the total denominator is the number of ONS LS members found in the 2001 Census. For category (5) the denominator is considered to be the total number of people found on the NHSCR at 2001 using the selection criteria for the ONS LS. **Table 1** shows the numbers found in each of the five groups.

Table 1 Categories examined, sample sizes and percentage of total*

Category	Numbers found	Per cent of total*
(1) Those present at 2001 Census[6] who also have an NHSCR record for the same area current at the 2001 Census date	516,758	95.7
(2) Those present at 2001 Census[6] but whose NHSCR record at census day in 2001 is in an area other than their census address	14,806	2.8
(3) Those present at 2001 Census[6],but whose NHSCR record had been cancelled prior to census day and no active record exists at 2001	4,434	0.8
3(a) Non armed forces	*3,155*	*0.6*
3(b) Armed forces	*1,279*	*0.2*
(4) Those present at 2001 Census[6] but not appearing until later on the NHSCR, if ever.	3,867	0.7
(5) Those 'absent' at 2001 Census[6] for whom an NHSCR record existed at census time*	88,961	14.9

* Note that the denominator for categories (1) to (4) is the total number of people found in the Longitudinal Study 2001 Census sample. The denominator for category (5) is the total number of people selected from the NHSCR using the selection criteria for the Longitudinal Study.

So, at the 2001 Census, over 95 per cent of people in the Longitudinal Study were resident in the same Health Authority area on the NHSCR as their address given in their census return. Of the

remainder, the majority were in a different Health Authority area (2.8 per cent), while 1.5 per cent were not on the NHSCR, with around half of these having never appeared there.

Category 5 is a particularly interesting group, but is quite difficult to interpret, as there are a number of reasons why there may be a GP record but no census record. These will include under-enumeration in the census and linkage issues in the LS, as well as those who have remained on the patient register but have left the country. **Table 2** gives a brief summary of what the data can tell us about this group.

Members can enter the ONS LS in three ways, through being found in a census, by being identified as an inward migrant into England and Wales in the NHSCR or through birth registration. Table 2 combines entry via census and from the NHSCR, but splits that category between whether the LS members were born in England and Wales or outside (migrants). Those that have entered through birth alone are then shown separately.

Table 2 Subsequent actions on NHSCR for those on register, but no census enumeration or census link, 2001-2006

	Entered ONS LS via census / NHSCR				Entered ONS LS		Total	
	E&W born	Per cent	Migrant	Per cent	via E&W birth	Per cent	Number	Per cent
No postings	20,282	63	12,889	44	17,156	63	50,327	57
Movement to:								
England or Wales	4,141	13	3,302	11	5,780	21	13,223	15
Scotland or Northern Ireland	186	1	177	1	163	1	526	1
Death	2,281	7	146	1	77	0	2,504	3
Cancelled	4,026	12	11,020	38	2,314	8	17,360	20
Armed Forces	27	0	8	0	100	0	135	0
Embarked	577	2	955	3	302	1	1,834	2
Live birth	400	1	629	2	1,421	5	2,450	3
Cancer registration	481	1	63	0	58	0	602	1
Total	**32,401**		**29,189**		**27,371**		**88,961**	

Nearly three fifths (57 per cent) of those found on the NHSCR but with no census form remain on the register with no further evidence of any activity. As stated previously, these people may have not responded to the 2001 Census, or there was an issue with linking their record to census data. It is also possible they may have left the country and not told their GP surgery (although list checking and cleaning should have identified this). One fifth were cancelled or embarked. Although the cancellation is after the census it is possible that some may have left before the census. The remaining fifth had activity which definitely indicated presence at some point after 2001. Patterns are broadly similar across the three groups of entry into the ONS LS identified in Table 2. However, the migrant group has a lower proportion with no activity and a higher proportion of cancellations. Further information on non-linkage and patient register data can be found in an article produced after the 2001 Census link[9]. The remainder of this article concentrates on those groups where we have census information, allowing a direct comparison between place at time of the 2001 Census and subsequent changes in NHS registration information captured by the ONS LS (categories (2) to(4) above).

Those present at 2001 Census but whose NHSCR record on 2001Census day is in an area other than their census address (category 2): what happened next?

Table 3 shows what subsequently happened to those who were at a different place on the NHSCR compared with census. The table shows that after six years over one half of this group did not end up being moved into the area in which the census found them. Of those, around one half had no other posting on the NHSCR.

Table 3 Next NHSCR event for those found in a different HA on NHSCR compared with the 2001 Census

	Number	Per cent
Moves them into their 2001 Census area	6,896	46.6
Moves them into another area	2,239	15.1
Shows an embarkation	42	0.3
Shows a cancellation	1,429	9.7
Died	142	1.0
Enlisted in the armed forces	76	0.5
Other	78	0.5
No post 2001 Census posting	3,904	26.4
TOTAL	**14,806**	**100**

The cases identified in the first row of Table 3 are a gross difference; however, it is quite possible that such differences may produce little net effect if, for example, for every person who is wrongly recorded on the patient register in Newcastle and recorded on the census in London there is a similar person (by age and sex) recorded on a patient register in London but recorded on the census in Newcastle.

Sample numbers are too small to examine this by individual health authority area, instead we have examined where the differences are by Government Office Region (GOR). Of the records examined just over half (51 per cent) were in the same GOR. **Figure 1** shows the gross numbers in a different GOR and the net numbers once cancelling has taken place to allow for compensating errors. Figure 1 shows that gross differences are concentrated in London and the South East, areas with large populations and with high levels of internal migration. Reassuringly, once gross differences are 'netted off' they become much smaller, although there is a clear geographical pattern with the East of England, South East and South West having net positive differences (that is there are slightly fewer people on the NHSCR in other regions who should be in those regions than there are people in the East of England, South East and South West regions who should be elsewhere), while more northern areas have net negative differences (that is there are slightly more people on the NHSCR in other regions who should be in those regions than there are people in those northern regions who should be elsewhere). It is particularly noteworthy that London and the South East have very small net differences despite having the largest gross differences. Of course, if there are very different effects at sub-GOR geography, then differences in the age-sex characteristics of the movers involved, or different time lags in subsequent moves between different areas, then these may still lead to quality issues. Although the sample numbers are too small to consider the timing of these moves between individual GORs, the overall average length of time for moves into and out of GORs can be calculated and are shown in **Table 4**. Note that this only refers to the average length (in months) of moves that have occurred by around six years after

the 2001 Census; the figures are, therefore, censored in that they don't include those who take more than six years to move.

Figure 1 Gross and net numbers where NHSCR has different GOR from census area, and patients subsequently move to census area

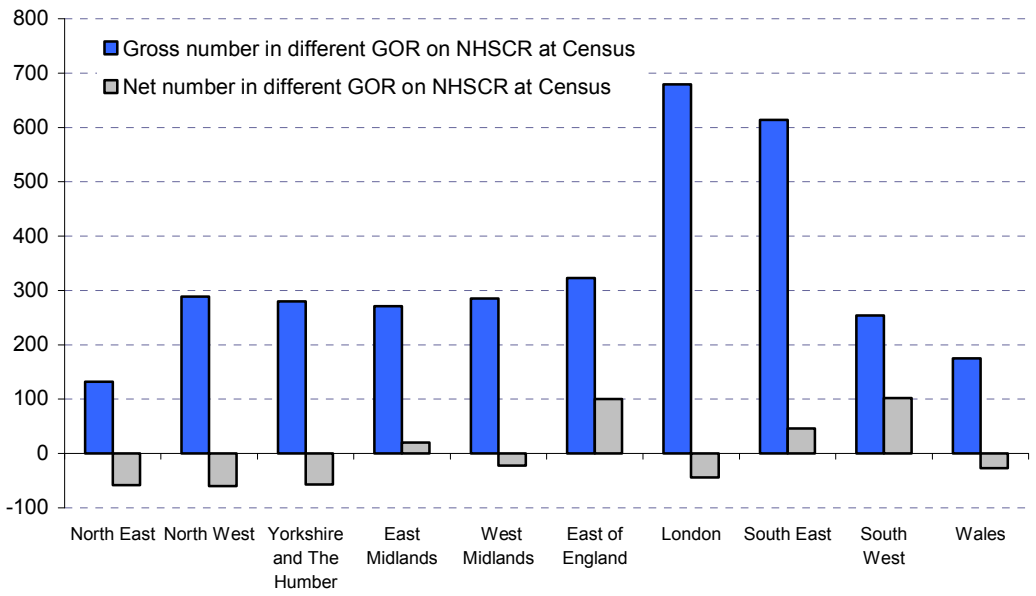


Table 4 Months taken to move from GOR on NHSCR to GOR in census for those who moved between the 2001 Census and the end of 2006

Government Office Region	Per cent moving from GOR	Per cent moving to GOR	Difference
North East	12.0	9.2	+2.8
North West	13.9	12.2	+1.7
Yorkshire and The Humber	13.5	13.3	+0.3
East Midlands	14.1	12.0	+2.0
West Midlands	13.9	12.9	+1.1
East of England	12.5	14.1	-1.6
London	14.1	15.4	-1.3
South East	13.4	13.3	+0.2
South West	12.6	13.3	-0.7
Wales	13.2	13.4	-0.2
Total	**13.5**	**13.5**	**-**

In general there are not large differences in time between moves into and out of an area. It is noticeable that it takes longer for people to be moved into London, the East of England, the South West and Wales than it takes for them to be moved out of those areas. This may relate to the type of people (younger, less likely to approach their GP), or it may relate to administrative differences. However these are averages, which are also censored, and ideally distributions would also need to be examined to draw any strong conclusions.

Those present at 2001 Census but whose NHSCR record had been cancelled prior to Census day and no active record exists at 2001 (category 3): what happened next?

Table 5 shows what subsequently happened to those who were cancelled on the NHSCR prior to census but who appeared in the 2001 Census.

Table 5 Next NHSCR event for those found in 2001 Census but with no active GP record

	Number	Per cent of total
All in Category		
Registration puts them into their 2001 Census area	1,829	41.3
Registration puts them into another area	768	17.3
Died	49	1.1
Enlisted in the armed forces	9	0.2
Shows an embarkation *		
Shows a cancellation *		
Other	92	2.1
No post 2001 Census posting	1,687	38.0
TOTAL	**4,434**	**100**
Of which: (a) Non Armed forces		
Moves them into their 2001 Census area	1,511	47.9
Moves them into another area	489	15.5
Died	46	1.5
Enlisted in the armed forces *		
Shows an embarkation *		
Shows a cancellation *		
Other	30	1.0
No post 2001 Census posting	1,079	34.2
TOTAL	**3,155**	**100**
(b) Armed Forces		
Moves them into their 2001 Census area	318	24.9
Moves them into another area	279	21.8
Died	3	0.2
Enlisted in the armed forces *		
Shows an embarkation *		
Shows a cancellation *		
Other	71	5.5
No post 2001 Census posting	608	47.5
TOTAL	**1,279**	**100**

* Categories have been collapsed for disclosure control

Around 41 per cent of these people reappear in the area in which the 2001 Census found them, although this increases to 48 per cent when excluding the armed forces. There were 38 per cent who had not reappeared over six years after census, although this reduces to 34 per cent when armed forces are removed. As these are people who have previously been selected as LS members and have been found in the census it might seem odd that they have not reappeared on the NHSCR six years later. There are also 17 per cent of this group who turn up in another area, the proportion being slightly lower for non-armed forces and understandably slightly higher for armed forces.

Those present at 2001 Census but not appearing until later on the NHSCR, if ever (Category 4): what happened next?

Table 6 shows those who had never been on the NHSCR at the time of the 2001 Census and what subsequently happened to them.

Table 6 Next NHSCR event for those found in 2001 Census, but with no previous GP record

Event	Number	Per cent of total
Moves them into their 2001 Census area	954	24.7
Moves them into another area	294	7.6
Died	25	0.6
Enlisted in the armed forces	0	0.0
Shows an embarkation *		
Shows a cancellation *		
Other	11	0.3
No post 2001 Census posting	2,583	66.7
TOTAL	**3,867**	**100**

* Categories have been collapsed for disclosure control

Around one quarter of those who are found in the census and have no previous GP record turn up in the area in which they were found in the 2001 Census. Two-thirds did not appear on the NHSCR in the six years following the census – this is equivalent to around 0.5 per cent of all linked records. Most of the remaining people in this category turn up in another area.

More detailed analysis by age and sex

It is possible to 'drill down' into the data provided to look at age/sex distributions and other characteristics from the census, for example student status. A limited amount of work has been carried out on this with the following tentative findings.

Table 7 demonstrates that there is a strong male bias for most of those who are either in a different place, are not on the NHSCR, or are on the register but absent from the 2001 Census. Given that there is evidence that census missed enumerating more young men than women, it suggests this may be a conservative estimate of the sex bias for those that are in the census[10]. The bias is likely to reflect well known issues of male engagement with GP services being both later and lower than for females.

Figure 2 shows the age patterns for the different groups for those present at 2001 Census.

Before considering the results in Figure 2 it should be remembered that by far the largest group is those for whom the 2001 Census and the NHSCR show the same place (see Table 1). It is clear however that all the other categories have a very much younger age distribution compared with that very large group.

The age distributions for those where the 2001 Census and the NHSCR disagree about the address, or where there is no NHSCR record prior to census, are very similar. Those where

NHSCR has been cancelled pre census, while still a younger distribution, is less peaked at younger ages, with a noticeably higher proportion of people in the 40-59 age group.

Table 7 Number of cases by sex and sex ratios of categories examined

Category	Men	Women	Sex ratio: men per 100 women
(1) Those present at 2001 Census who also have an NHSCR record for the same area current at the 2001 Census date	247,135	269,623	91.7
(2) Those present at 2001 Census but whose NHSCR record on census day in 2001 is in an area other than their census address	9,015	5,791	155.7
(3) Those present at 2001 Census but whose NHSCR record had been cancelled prior to census day and no active record exists at 2001	2,765	1,669	165.6
3(a) Non Armed forces	1,713	1,442	118.8
3(b) Armed forces	1,052	227	463.4
(4) Those present at 2001 Census but not appearing until later on the NHSCR, if ever.	2,346	1,521	154.2
(5) Those' absent' at 2001 Census for whom an NHSCR record existed at census time	51,570	37,391	137.9

Figure 2 Age distribution of different categories of those present at the 2001 Census

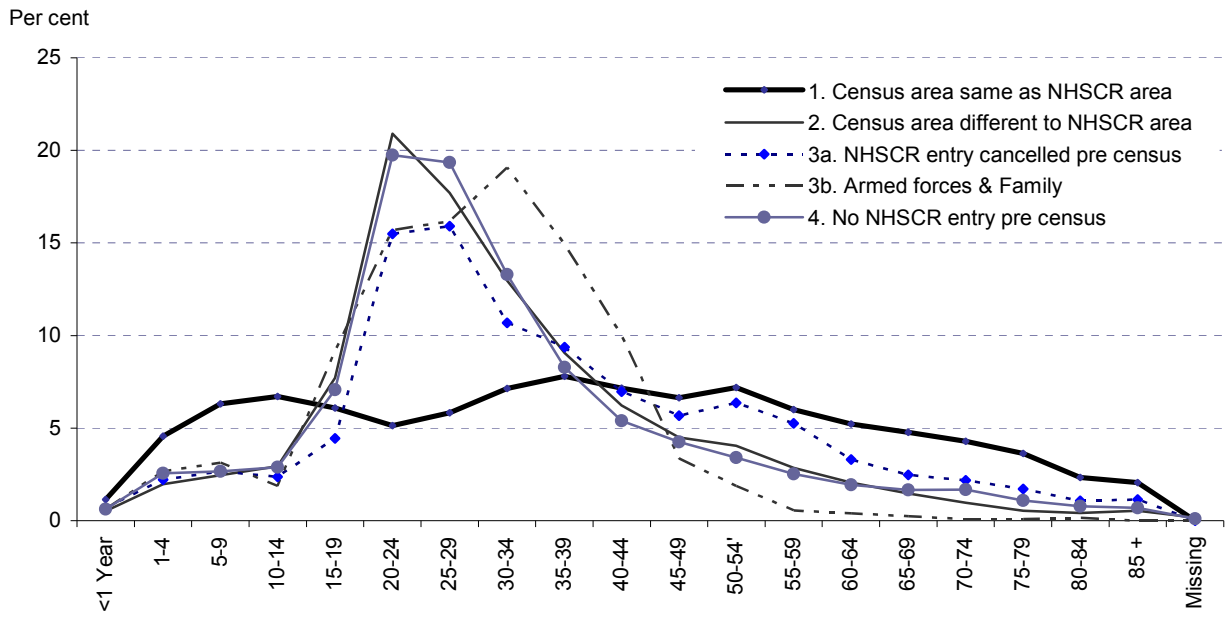
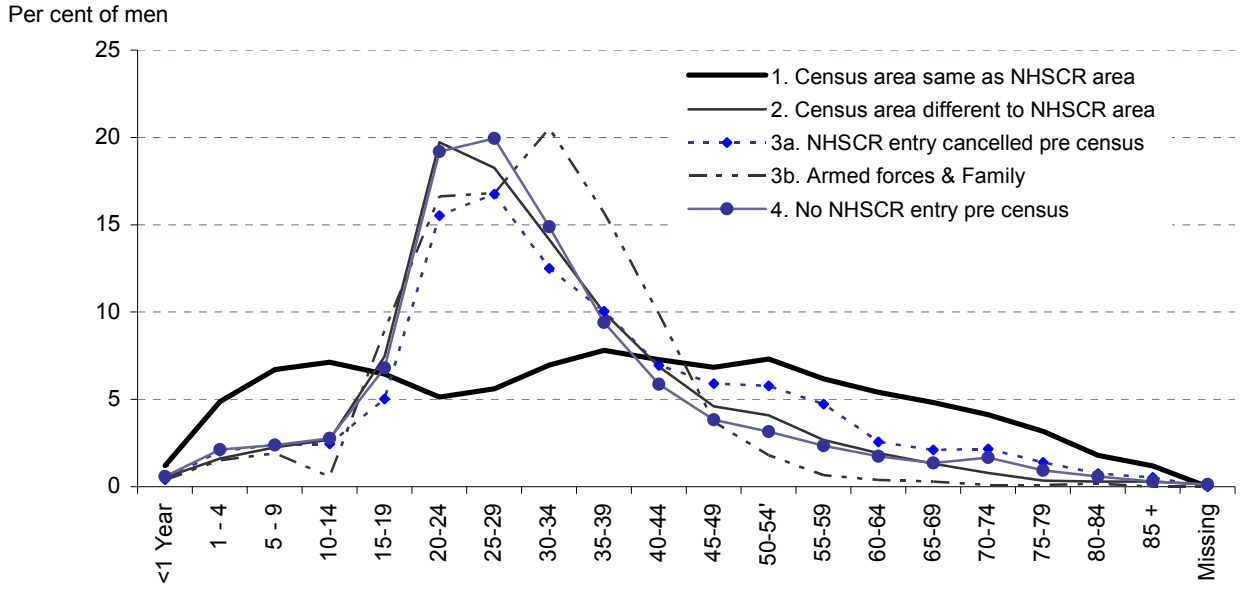


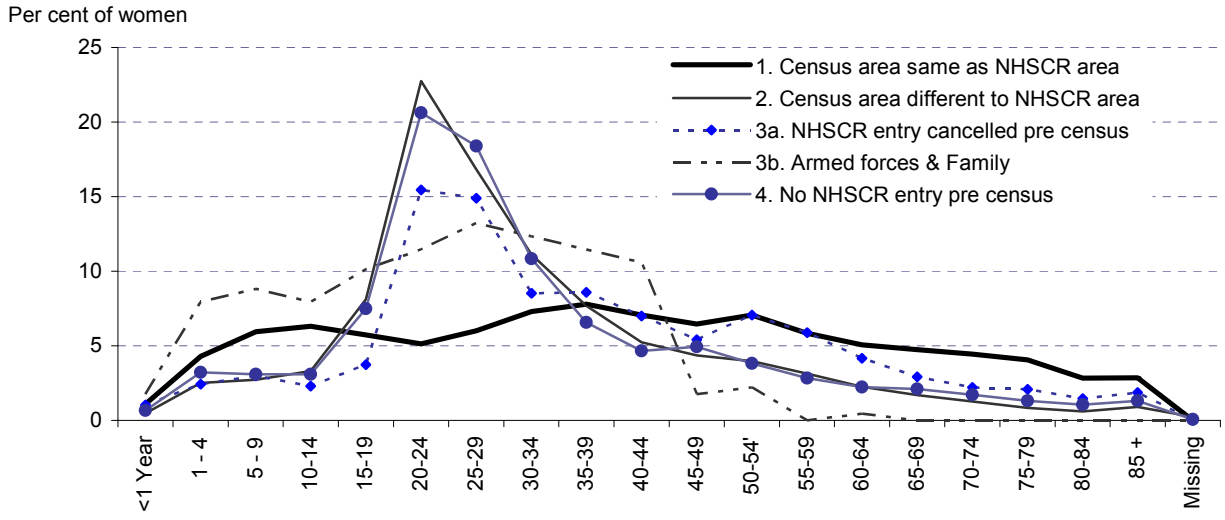
Figure 3 shows the data in Figure 2 separately for males and females. The patterns for males and females are very similar, although the female 'NHSCR entry cancelled pre-census' group has a distribution even closer to the overall population distribution than the male distribution.

Figure 3 Age distribution of different categories of those present at 2001 Census, male and female

(a) Males



(b) Females



Students

A further subdivision of the data was carried out to see if there was any particular effect from those who were recorded as students in the 2001 Census. **Table 8** shows that, broadly, there were similar proportions of students in the groups investigated. This is perhaps surprising given that all the groups where the 2001 Census does not give the same address as NHSCR are much younger on average.

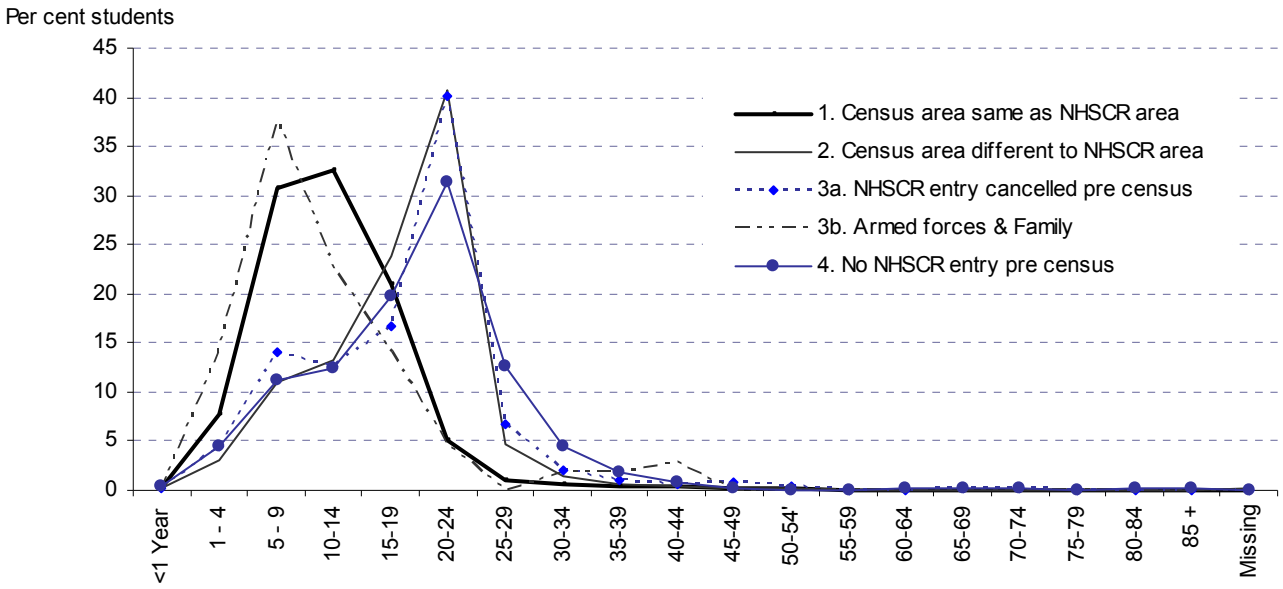
Table 8 Those present at 2001 Census by NHSCR location, and whether student or not

	Students	Per cent	Non-students	Per cent	Total
1. Census area same as NHSCR area	106,254	20.6	410,504	79.4	516,758
2. Census area different to NHSCR area	3,273	22.1	11,533	77.9	14,806
3a. NHSCR record cancelled pre census	598	18.9	2,557	81.1	3,155
3b. Armed forces & Family	106	8.3	1,173	91.7	1,279
4. No NHSCR record pre census	907	23.5	2,960	76.5	3,867
Total	**111,138**	**20.6**	**428,727**	**79.4**	**539,865**

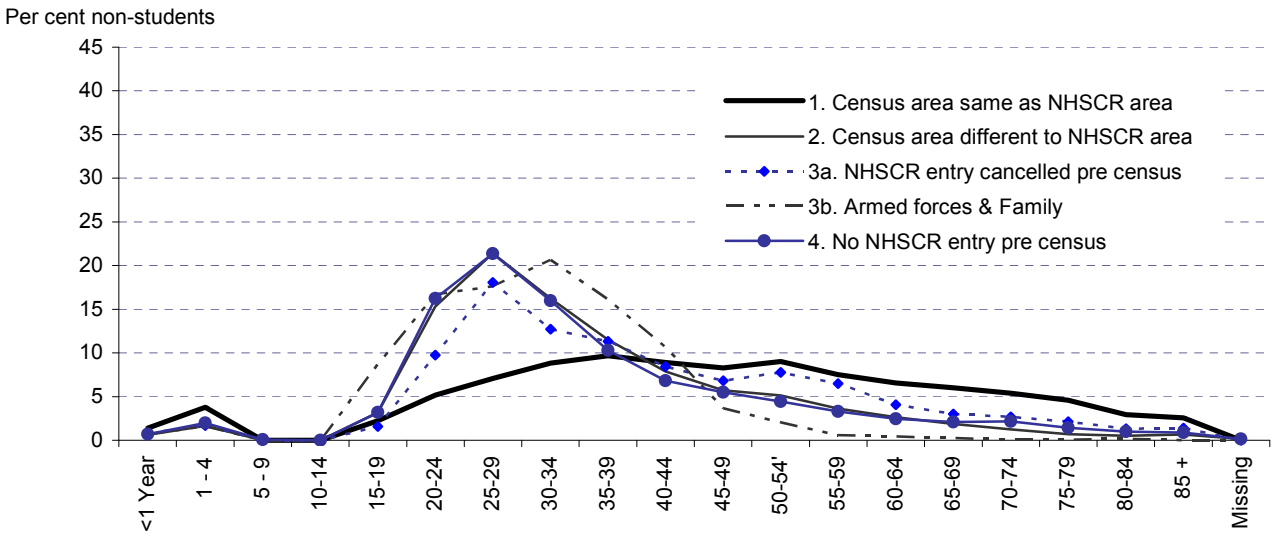
The age distributions show that students are much younger than the non-students in each of the groups. **Figure 4** shows the age distributions for students and non-students, the great majority of students being aged under 25; that students are likely to be particularly important for differences in the 20-24 age group (between ages 5 and 19 the majority of the population are students, but are also dependent children). However, it is possible that those who were ever students are affecting the higher age ranges. Further work may be possible to analyse this group using educational qualification.

Figure 4 Age distribution of different categories of those present at 2001 Census, students and non-students

(a) Students



(b) Non students



Births and deaths

The ONS Longitudinal Study also contains registrations of births and deaths. Births and deaths are used in the construction of population estimates, while those who are born or have died are added to/removed from the NHSCR. Comparisons were made only where first NHSCR entry (for births) and the last NHSCR entry prior to death was in England and Wales. For births the comparison was made between the residential address of the mother given at the registration of the birth and the first posting (health authority area of registration) of the child born, for children selected into the

ONS LS. For deaths, the comparison was with the residential address given for the deceased at death registration, together with the final posting on the NHSCR for the death. Only 1.7 per cent of births (564 of 33,298) and 1.9 per cent of deaths (579 of 30,484) were found to have a different health authority area when the sources were compared; for 40 per cent of the deaths with a different area there was evidence that the person had moved within the previous calendar year: it could be hypothesised that these are moves to be closer to relatives, perhaps to live with them or into sheltered accommodation, care homes and hospices. It is reassuring that most deaths are within the same health authority area in both sources, and also that the patient register picks up these final moves. However, in the small number of cases where there is a different address and evidence of a recent move, there is an issue for population estimation. These final moves will be included in the estimation of internal migration, while the deaths will be subtracted from the previous health authority area.

Length of time to recording of change

For those who are recorded in a different place on the patient register than where they are found in the 2001 Census, but then move, it is possible to look at how long it then takes before that move is recorded on the register. The first group (category 2) are those who were subsequently recorded as having moved to their census area.

Figure 5 shows the patterns for males and females. Note that this is only those who have moved; there are more males than females enumerated in a different area at time of census, and, similarly, there are more males in the category of moving (4,205 males and 2,845 females). As a proportion of the total, those who have moved is similar across the sexes (46 per cent males and 48 per cent females). Within one month, one in 10 males and around one in seven females moved to their census area. Of all the women who moved, half moved by the sixth month following the census; for men this was in the tenth month. Of course, these are marginal differences in terms of calculating population estimates, but it does suggest that there is a material difference in the speed of re-registering with a GP for men compared with women.

Figure 5 Length of time taken to move to place of census enumeration in NHSCR

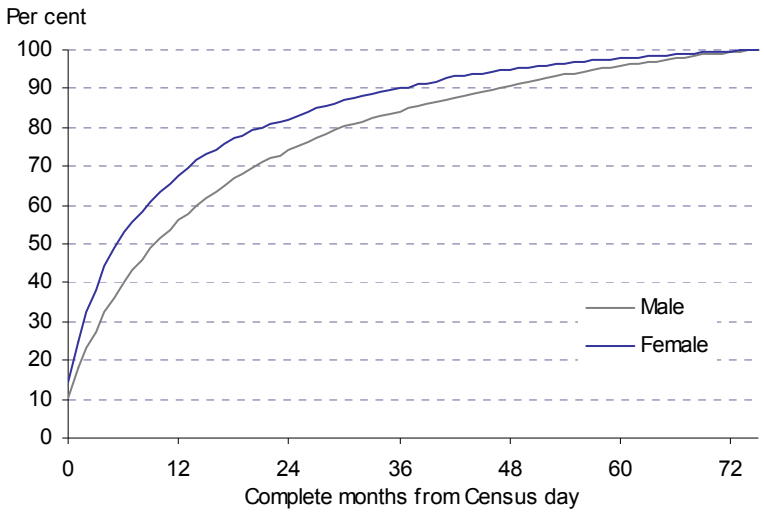


Figure 6 shows a similar analysis for those with no entry on the patient register prior to census. Only 23 per cent of men and 27 per cent of women end up with an NHSCR entry in the census area. For those who do, there is still a difference between men and women in the timing, but the differences are smaller than in figure 4, and the movement is slightly faster.

Figure 6 Length of time taken to appear at place of census enumeration where no entry on patient register at time of 2001 Census

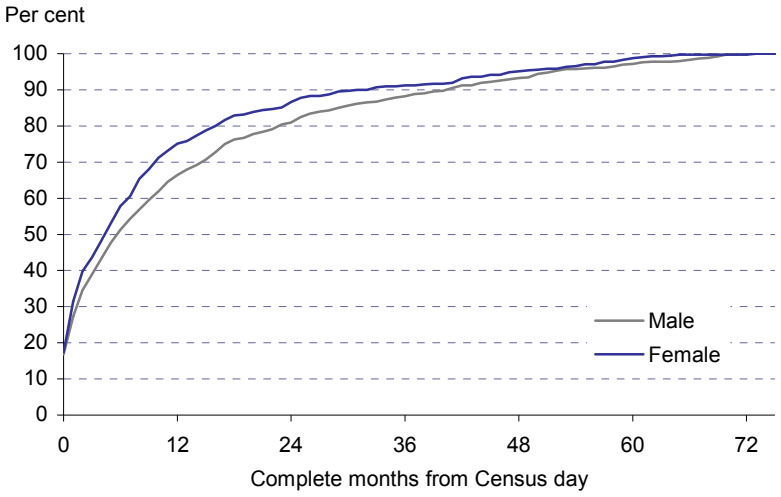
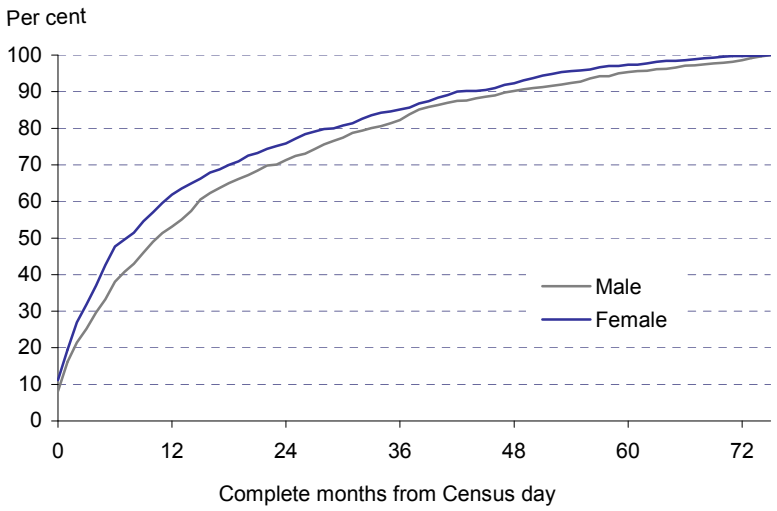


Figure 7 Length of time taken to appear on patient register at place of census enumeration after census where previously cancelled on the NHSCR



Finally, **Figure 7** shows the patterns for people reappearing on the NHSCR after the census, having been cancelled on the register prior to census. Around 47 per cent of men and 49 per cent of women re-appear. Again women re-appear more quickly than men. Some caution is required

here as this analysis does not consider how long before census the cancellation took place. Also the cancellation may have been legitimate if the person had emigrated at some point.

Discussion

This article has shown the richness of administrative data that can be produced when it is compared with census data in a longitudinal analysis. The analysis provides reassurance that, generally, the patient register is a good quality administrative source, and in most cases matches with the information collected at census. Even where there are lags in the movement of people recorded by the NHSCR the proportion is not large and there is a lot of net cancelling, at least at the GOR level. Where there are quality issues, the ONS LS is able to quantify the extent of those issues and differentiate them by age, sex and population subtype. This analysis has looked at the sub-type of students; it shows how other data sources which identify population and place, if they were included in the ONS LS, might be analysed to show their quality with regard to place. This could be particularly important as sometimes claims of deficiency are made against administrative data, often because there is insufficient incentive for users to update their address.

As patient register data is currently an input into making population estimates, which are used for resource allocation, it is important to understand their strengths and limitations. Increasingly, however, there is pressure to move away from the traditional systems of taking censuses and then creating subsequent population estimates using cohort component methods, rather than using the administrative sources themselves to produce the population estimates. Such a move gives even greater importance to understanding the quality of the administrative sources. This article points the way to how other data sources could be examined, if they could be linked into the ONS LS.

This article has not covered all the potential of analyses, possible using the ONS LS, to look at patient register data. As well as further work on students, other possible work could be a more detailed analysis of the information on migration recorded in the census from the question about the respondent's address one year ago. Examining this would assist in interpreting the timing data, since the sub-group that has moved before the census, but still have an address one year ago on the NHSCR, can be followed up to see if and when that move was recorded. A direct comparison of the moves that took place between May 2000 and May 2001 (allowing a one month lag for moves to take place) on the NHSCR could also be compared with those that identified themselves as moving internally to supplement macro comparisons already carried out[3].

Conclusion

The proportion of the population that completes a census form but that never appears on the NHSCR is small (estimated at 0.5 per cent of those enumerated in 2001 Census). While there is an issue around *where* census counts people, and *where* those people currently are in the NHSCR, the issue does appear to be small. However, it may need to be considered when initially rolling forward from 2011 Census to the mid-year estimate to prevent double counting of movers.

This work provides further evidence that men are more likely to be mis-recorded in GP registers compared to women. The proportion of students is slightly higher in the groups who are registered in a different area from census or are found in census, but not on the NHSCR. The students in

these groups are younger than the non-students. However, overall, students do not dominate the numbers in those categories where 2001 Census and NHSCR do not match.

The issues around births and deaths seem very small, being less than two per cent of the total number of events, which itself is a relatively small proportion compared to the total population.

This work essentially relates to a 'snap shot' in 2001. It cannot answer the question: *are things better or worse now?* In particular, since 2001, there has been a growth in numbers of international migrants; and this work does not take that growth into account. Once the 2011 Census results are linked to the ONS LS, and time has elapsed to allow for subsequent information to accrue, this work can be updated.

Key findings

- Overall, 95.7 per cent of ONS LS members enumerated at the 2001 Census resided in the same area as recorded on the NHSCR data.
- The proportion enumerated in the census that never appear in NHSCR registration is small, estimated at 0.5 per cent of those enumerated in 2001 Census.
- This work provides evidence, consistent with other analyses, that men are more likely to be mis-recorded in NHSCR registration than women.
- Where there are lags in movements in the NHSCR system there is evidence that once compensating differences are taken into account the net effect by Government Office Region is small.
- The proportion of students is only slightly higher in the groups who are at a different address from census, or are found in the 2001 Census but not on NHSCR. However, this work only identifies those that are currently students; it is unable at present to identify those who were ever a student.
- The article also compares registrations of births and deaths in the period 2001-06. Differences in geography around births and deaths look very small, being less than two per cent of the total number of events, which in themselves are a relatively small proportion when compared with the total population.

Acknowledgements

The authors wish to acknowledge the helpful comments of anonymous reviewers and the assistance of the Population Trends editorial and production team.

References

1 Office for National Statistics (2002) 'Using Patient Registers to estimate Internal Migration – Customer Guidance Notes'. Migration Statistics Unit, ONS, August 2002. Available at: www.statistics.gov.uk/statbase/EXPODATA%5Ccommentary%5CGuidance%20Notes.pdf

2 Bates, AG (2004) 'Small Area Population Estimates project: data quality of administrative datasets'. *Population Trends* 116: 11-17.

3 Available at: www.statistics.gov.uk/downloads/theme_population/Revising_Internal_Migration_Estimates.pdf

4 Hattersley, L and Creeser, R (1995) *Longitudinal study 1971-1991: History, organisation and quality of data*. Office for National Statistics Series LS7. HMSO: London.

5 www.ons.gov.uk/about/who-we-are/our-services/medical-research/about-us/national-health-service-central-register--nhscr-/index.html

6 www.statistics.gov.uk/geography/england_health.asp

7 'Present at 2001 Census' in this context means that there is a census record linked to the LS study.

8 Those that have left England and Wales and registered with a doctor in Scotland or Northern Ireland are identified on CHRIS.

9 Blackwell, L, Lynch, K, Smith, J and Goldblatt, P (2003) *Longitudinal Study 1971–2001: Completeness of Census Linkage*. Series LS 10, Office for National Statistics and Centre for Longitudinal Studies, Institute of Education, University of London, London.

10 Abbott, O, Diamond, I and Jackson, N (2003) 'Key Issues in the Quality Assurance of the One Number Census'. *Population Trends* 113: 11-19.