Rees P, Clark S, Wohland P, Lomax N & Norman P (2016) Using the 2001 and 2011 censuses to reconcile ethnic group estimates and components for the intervening decade. Chapter 21 in Stillwell J and Duke-Williams O (eds.) Unlocking the UK 2011 Census: Handbook of Census Resources, Methods and Applications. Routledge, Abingdon.

Draft Pre-Proofs

Chapter 21

Using the 2001 and 2011 Censuses to Reconcile Ethnic Group Estimates and Components for the Intervening Decade for English Local Authority Districts

Philip Rees, Stephen Clark, Pia Wohland, Nikolas Lomax and Paul Norman

Abstract

This chapter describes the creation of new estimates of ethnic populations and components of change in local authority districts (LADs) in England for years between the 2001 and 2011 Censuses. Information on ethnic populations by age and gender is provided in censuses. In between censuses, information on ethnic population change is scarce. To fill the gap we used data from the two censuses with reconciled total population and component estimates published by the Office of National Statistics. This chapter outlines the sequence of steps used to produce a ten-year time series. These reconciled population and component estimates provide a firmer foundation for ethnic-specific projections than hitherto available. The role of the census in this work is vital.

21.1 Introduction

Periodic censuses, annual population estimates and projections all provide important data for monitoring the state of the UK's population, where the aim is to count, estimate or project everybody. These datasets can be supplemented by survey data and administrative records which cover sub-sets of the population. Integration of these elements is currently being explored (ONS, 2015). UK census tables provide detailed population data down to small area scale (e.g. output areas); the mid-year official estimates provide populations by age and sex for local authority districts (LADs) and sub-LAD areas (e.g. super output areas, wards); official projections provide population tables by age and sex and households for LADs for the 25 years. Compared with full censuses, the estimates and projections are very 'informationlite'. Should we be interested in reliable information beyond age and sex for the whole population, reliance must be placed on the census.

Population estimates and projections have been developed for various attributes beyond just age and sex. Ethnicity has been estimated at LAD scale by the Office for National Statistics (ONS, 2011) and the ethnic composition of the population has been projected at LAD scale in England by Wohland et al. (2010) and Rees et al. (2011; 2012) and for London Boroughs by the Greater London Authority (GLA, 2015). Forecasts of health status combining census, survey and population projection data for older people in English LADs have been generated by Clark (2015).

UK population estimates and projections are based on rolling forward populations from census counts using estimates of the components of change based on register and administrative data in the short term and assumptions about future behaviour of the rates in the long term. Evaluation of LAD population estimates by ethnicity rolled forward from the 2001 Census to mid-2009 (ONS, 2011) and the ETHPOP projections by ethnicity from mid-2001 to mid-2011 revealed serious departures from the results of the 2011 Census (Rees et al., 2016a).

The aim of this chapter is to explain how we estimated the populations and components of change by ethnicity, age and sex for the LADs of England together with Wales, treated as a single zone. Because population level information (rather than sample based or proxy administrative information) on ethnicity is virtually absent in UK demographic statistics, the methods used here rely on the 'book ends' of the 2001 and 2011 Censuses. The ethnic group estimates are constrained so that they sum to the official mid-year population estimates and the mid-year to mid-year components of change. The time series of component rates by ethnicity provides the basis for setting assumptions for projections with greater intelligence.

The chapter is focussed on methods used; empirical results will be discussed elsewhere. Section 21.2 reviews previous work on estimating ethnic populations and components of change between censuses. Section 21.3 describes the adjustments needed to harmonize populations between two censuses by geography, time and ethnicity. Section 21.4 presents the age-time frameworks required in the estimation. Section 21.5 sets out the flows charts and equations used to interpolate mid-year populations and populations at risk for two types of cohort while section 21.6 explains the methods used to estimate the components of change. We bring together the populations and components in section 21.7 in demographic account tables, implementing a final adjustment to ensure that demographic inputs equal outputs. The final section summarises and evaluates the methodology.

21.2 Estimating ethnic populations and components between censuses

One technique, employed since censuses have been taken and where good registers of births and deaths exist, is to infer net migration for different populations by subtracting natural change from population change between censuses. When the populations are disaggregated by age, then the calculation simplifies into a subtraction of deaths from population change. This technique is best applied to age bands with intervals equal to the inter-census time interval. It has not been used to estimate the components of change for populations classified by ethnicity.

Where estimates of inter-census populations by ethnicity are made, a cohort-component method is used. For example, the ONS produces roll forward ethnic population estimates for LADs in England (Large and Ghosh, 2006a; 2006b; ONS, 2011). The method depends on estimation of the components of change (births, deaths, internal and international migrations) for each year intervening between two censuses. These estimates assumed that mortality rates were the same for all ethnicities, concealing important variation (Wohland et al., 2015) and comparison with estimates based on Annual Population Survey (APS) data suggested considerable differences.

Bryant and Graham (2013) recognise that both population estimates and projections can drift away from the later ‘true’ figures. They construct a method for subnational population estimation using a formal Bayesian framework. This enables them to develop credible intervals around estimates and demonstrate how uncertainty can be controlled for estimates between censuses. This uncertainty, however, can escalate without the constraint of a second census. At the heart of the estimation is a demographic account, providing a complete description of the population system. The relationship between the account and data is described by an ‘observation’ model which involves simulation carried out using Markov Chain-

Monte Carlo methods (Bryant and Graham, 2013). We considered but did not use the Bryant-Graham methods, because of the computational challenge of processing 12 ethnic groups and 324 LADs in England compared with 4 ethnic groups and 67 territorial authorities in New Zealand.

21.3 Getting the census populations into shape

In order to use population data from the UK 2001 and 2011 Censuses to estimate populations and component rates by ethnicity for mid-year intervals between mid-2001 and mid-2011, it is necessary to harmonize key variable definitions for ethnic groups and LADs. Harmonization is needed because the ethnic classifications differ between censuses and between the different 'census' countries: England and Wales, Scotland and Northern Ireland. The ONS harmonized classification uses 10 groups but several features are problematic. The Gypsy/Traveller/Irish Traveller category is very small and any component estimates are likely to be poor. The absorption of the White Other group into a general White category means that significant new population flows in the decade cannot be tracked: the immigration from Eastern European states that joined the European Union (EU) in May 2004 and the immigration from Southern European states with high unemployment. Finally, in the ONS harmonized classification, the three Black groups, distinguished in both censuses with different ethnic origins and demographic profiles, are merged into one. So, in the classification used in our analysis we move the Gypsy/Traveller/Irish Traveller population into the White British and Irish group and retain a separate White Other group. We accept the ONS merger of the different mixed or multiple ethnicities into one Mixed group. However, we retain three distinct Black groups. None of these decisions causes problems for LADs in England.

Harmonization of geographical areas is essential in any analysis of population change. The principle we adopted was to employ the most recent definitions of LADs (lowest tier) in each home country of the UK. The current analysis is focussed on English LADs, which were re-organised in 2009 through the creation of new unitary authorities (UAs) in place of previous shire counties and districts. Where this occurred, 2001 Census shire districts were aggregated to their successor UAs. We used two mergers of LADs with small populations into larger neighbours from previous analysis: the Isles of Scilly were merged with the Cornwall UA and the City of London with the City of Westminster. In this analysis, we used 324 English LADs and Wales is treated as a single zone.

21.4 The time and age-time frameworks for inter-census estimation

We need a clear framework for reconciling populations by age and ethnicity at the two censuses with the intervening components of change. By reconciliation we mean that the end population defined by the second census is obtained by inputting the components of change for the decade. To avoid complexity, we adjust the LAD populations at the census by ethnicity, age and sex, counted on 29 April 2001 or 27 March 2011, to agree with the following mid-year LAD population by age and sex. UK demographic statistics use mid-years as the time points at which the population is estimated or projected. This means we have an exact 10-year interval between the mid-year estimates directly informed by the first and second censuses. We label these adjusted counts 'Census-Based Book End' (CBBE) populations and CBBE components.

There are two parts to this framework: an identification of the time points and intervals for which data are available; and a specification of the relationship between age and time. The columns in Figure 21.1 indicate the time intervals and points for which demographic data are estimated, either for calendar years or for mid-year to mid-year intervals. The rows represent the four components of change. Fertility and mortality are assigned two rows because the available input data refer to calendar years while the data needed for the reconciliation exercise refer to mid-year to mid-year intervals. The mid-year to mid-year demographic flows are estimated from the calendar-year data for births and deaths. The internal migration data are published for mid-year to mid-year intervals, while international migration data are provided for a variety of annual periods and are converted into mid-year to mid-year estimates.

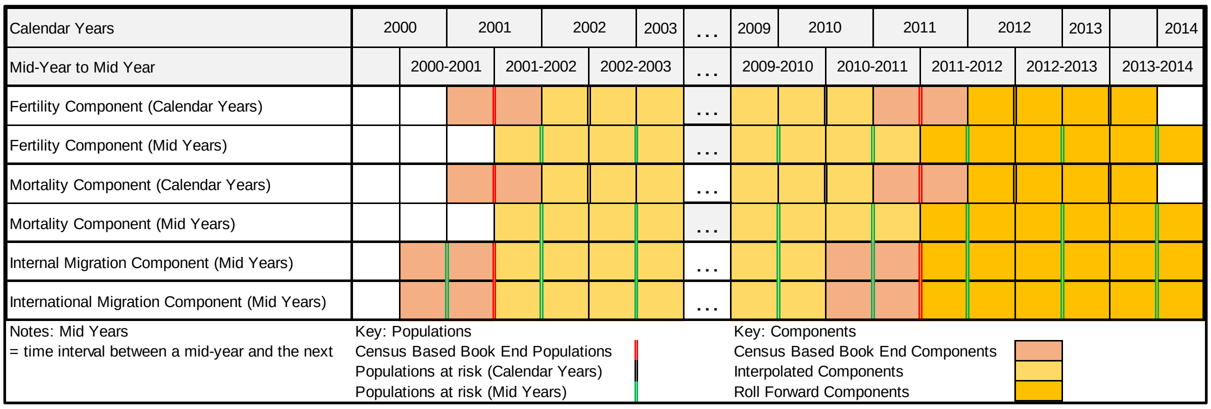


Figure 21.1 A time framework for estimation and projection of populations and components by ethnicity

Figure 21.1 represents populations counted or estimated for a point in time as vertical lines in different colours. It shows components (events or transitions that change the population) as horizontal bars, shaded in different colours. The CBBE adjusted counts are represented as red double lines, positioned at mid-2001 and mid-2011. The time intervals for component flows and rates extend from one mid-year to the next. However, often component flows are reported only for calendar years. Data from two successive years need to be combined, usually through averaging. CBBE components are identified in a pink shade. We sum the initial estimates of the LAD components by ethnicity, age and sex and then adjust these totals to agree with the ONS LAD components.

Mid-year LAD populations by ethnicity, age and sex are shown as vertical lines in the figure, from which populations at risk (PAR) are computed. The PAR for fertility and mortality rates are the mid-year populations of the calendar year, shown as vertical black double lines. The PAR for internal and international migration are an average of mid-year populations in the two years making up the mid-year to mid-year intervals and are shown as vertical green double lines. The PAR are multiplied by the component rates to yield first estimates of ethnic specific flows, which are adjusted to the all-person estimates produced by ONS. The migration statistics, which refer to the year prior to the census, are used in combination with migration statistics for the mid-year to mid-year interval preceding the CBBE population. For international migration the interpolation is carried out on flows rather than rates because we use assumptions about future flows rather than rates in subsequent projections.

The second part of the framework is provided by the age-time diagram, a graph in which age is plotted on the vertical axis and time on the horizontal (Figure 21.2). The age band widths and time intervals in an age-time diagram must be equal in a roll-forward population estimation or projection model. Figure 21.2A sets out the standard case and Figure 21.2B the case for new-borns. The vertical axis is marked into equal sized bands of single years of age and labelled with age at last birthday. There are 10 annual time (mid-year to mid-year) intervals on the horizontal axis and 20 ages (at last birthday) shown on the vertical axis. Horizontal bands refer to the same age over time intervals that change, while vertical bands refer to the same time intervals over ages that change. Diagonal bands in the graph refer to persons belonging to the same birth cohort. Each parallelogram in the graph refers to a period-cohort which tracks the changes in a cohort population in an annual time interval.

In Figure 21.2A, selected lines have been highlighted by colour, marking off a block of period-cohort spaces. The block marked out by bolded black lines shows the progression of persons aged 0-9 at mid-year 2001, the starting CBBE population which is highlighted in red, to being aged 10-19 at mid-2011, the finishing CBBE population, highlighted in green. The brown lines in the diagram delimit the age-time trajectories of a single cohort, born between mid-year 1995 and mid-year 1996. Members of this birth cohort are aged 5 at mid-year 2001 and aged 15 at mid-year 2011. What we need to do is to find the set of period-cohort component flows which, over the ten years (mid-year to mid-year intervals), change the starting population into the finishing population. The standard age-time framework applies to ages 0-100+ at mid-year 2001. Babies born each year and require special treatment (Figure 21.2B): births replace the 2001 CBBE population as the starting book-end population and the estimates are made over time intervals which shrink from 9.5 years to 0.5 of the 2001-11 decade.

21.5 The estimation of mid-year populations and populations at risk

Before proceeding with the estimates of mid-year populations, the information available on ethnicity for mid-2002 to mid-2010 was reviewed for use in the estimation. The candidates for proxy populations were: (i) the Trend and/or the Emigration Rate projections for 2001 to 2011 from the 2001 Census-based ETHPOP projections (Wohland et al., 2010; Rees et al., 2011; 2012); (ii) the Population Estimates by Ethnic Group (PEEG; ONS, 2011); (iii) Annual Population Survey (APS) estimates (ONS, 2016); and (iv) linear interpolation between CBBE populations. Following evaluation of each alternative (Rees et al., 2016), linear interpolation between CBBE populations for LADs by gender, age and ethnicity was implemented.

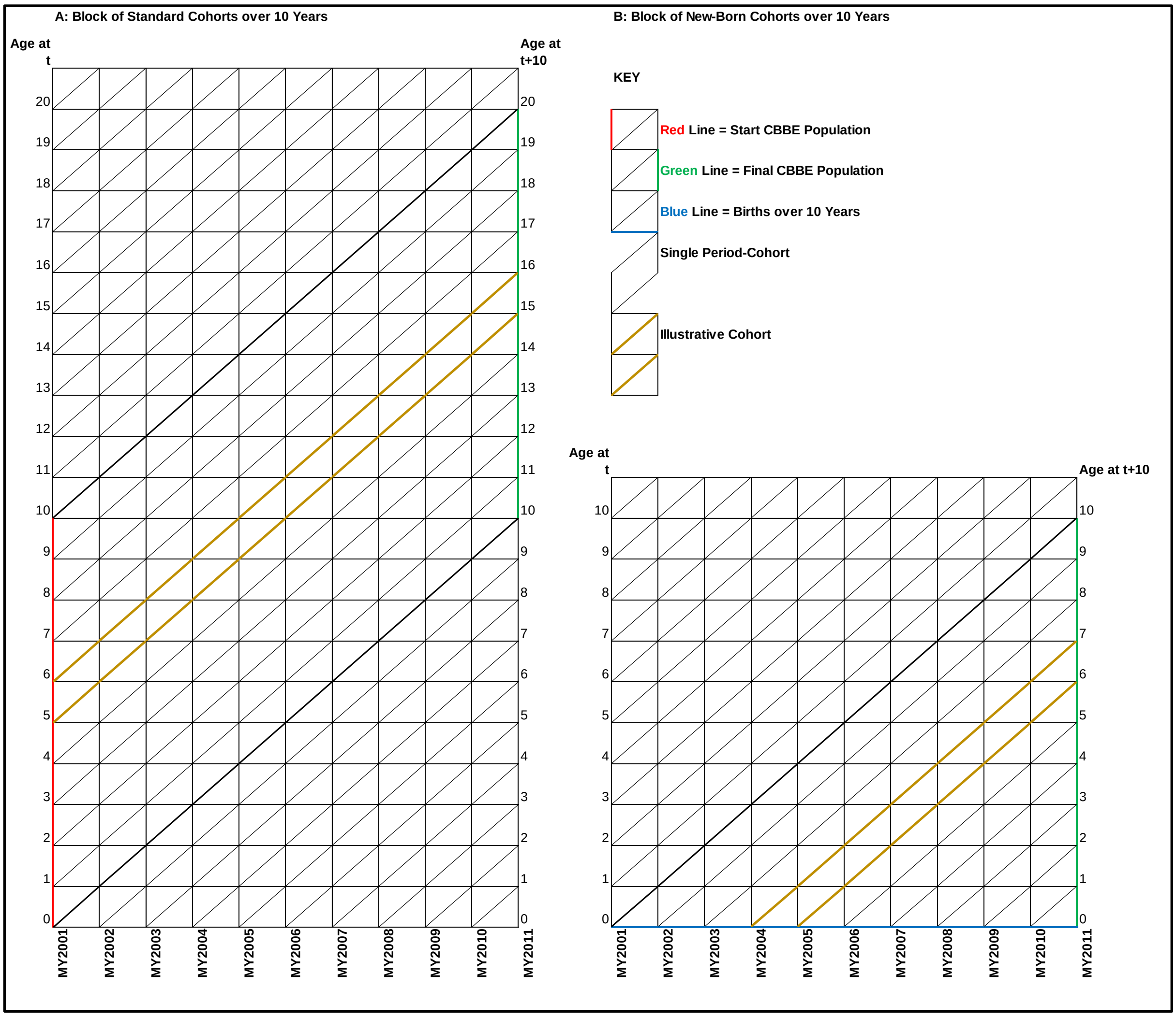


Figure 21.2 Age-time diagram for standard and new-born period-cohorts over 10 years

Figure 21.3 sets out the scheme used in computing the ethnic specific mid-year populations, in this case for 2000-01 and 2010-11. The top row of boxes indicate the inputs to the population estimation. The box shaded in blue refers to the census populations aggregated into our 12 harmonized ethnic groups. The boxes shaded in grey contain the mid-year population estimates by gender and age, reconciled by ONS between the 2001 and 2011 Censuses (ONS, 2013). The left-most box refers to the mid-2000 population reconciled to the 1991 and 2001 Census populations. The second row contains the CBBE populations in pink shaded boxes for 2001. The mid-year populations by gender, age and ethnicity are linearly interpolated between the CBBE populations in 2001 and 2011 as follows:

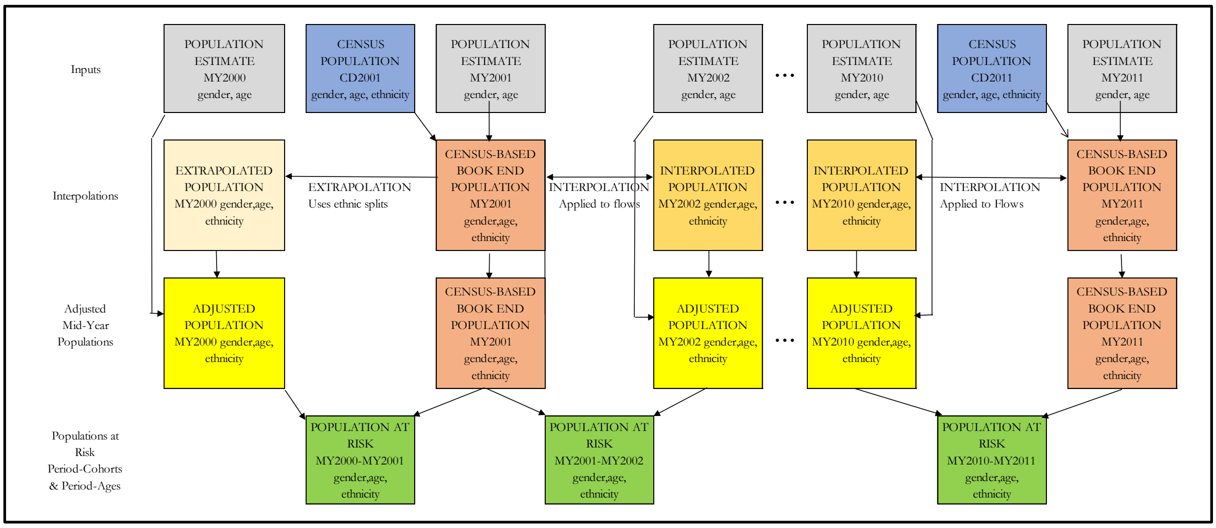
$$P_{g,x+y,e}^{i,2001+y} = P_{g,x,e}^{i,2001} + \left(\frac{y}{10}\right) \times \left[P_{g,x+10,e}^{i,2011} - P_{g,x,e}^{i,2001}\right] \quad (21.1).$$

where P stands for population, i for LAD, g for gender, x for age (0, 100+) and e for ethnicity; y (= 1, 9) is the mid-year number counting forward from year zero (mid-year 2001). A tenth of the difference between the 2001 and 2011 CBBE populations in the same cohort, 10 years of age apart, is multiplied by the year index and added to the mid-2001 population. This equation applies to the cohorts aged 0-100+ at mid-year 2001. For those ages where the population is born during the period, we implement a scheme for people starting life in the interval between the CBBEs for birth years MY2001-MY2002 to MY2010-MY2011:

$$P_{g,x,e}^{i,2001+(10-y)+x} = B_{g,e}^{i,2001+(10-y)} + \left(\frac{x+0.5}{y+0.5}\right) \times \left[P_{g,y,e}^{i,2011} - B_{g,e}^{i,2001+(10-y)}\right]$$

$$\forall\, y \in [9,0], x \in [0, y-1] \quad (21.2)$$

The equation is shown using a nested ‘loop’ where the outer loop, y, is the age at MY2010-MY2011 and the inner loop, x, is the age being estimated from the new-born cohort to the year before the target age in MY2010-MY2011, i.e. (y-1). In order to interpolate between births in each year and the CBBE population for an age which is linked by cohort survival to births, we need to implement the fertility component estimation for the maternal age range 15 to 49 (discussed later in section 21.6).



Notes: MY = mid-year (30 June/1 July), CD = Census Date (29 April 2001)

Figure 21.3 Scheme for estimating ethnic-specific populations and PAR, mid-year 2001 to mid-year 2011

An additional computation is needed for the mid-2000 population where we extrapolate the ethnic composition of the 2001 CBBE population backwards using ethnic shares of each LAD, gender and age population group, adjusting to ensure that shares are non-negative and sum to 100% as follows:

$$P_{g,x,*}^{i,y} = \sum_e P_{g,x,e}^{i,y} \quad (21.3)$$

$$S_{g,x,e}^{i,y} = 100 \times [P_{g,x,e}^{i,y} / P_{g,x,*}^{i,y}] \quad (21.4)$$

$$S_{g,x-1,e}^{i,2000} = S_{g,x,e}^{i,2001} - \left(\frac{1}{10}\right) \times [S_{g,x+10,e}^{i,2011} - S_{g,x,e}^{i,2001}] \quad (21.5)$$

$$S_{g,x-1,e}^{i,2000\prime} = S_{g,x-1,e}^{i,2000} \times [100 / \sum_e S_{g,x-1,e}^{i,2000}] \quad (21.6)$$

$$P_{g,x-1,e}^{i,2000} = P_{g,x-1,*}^{i,2000} \times [S_{g,x-1,e}^{i,2000\prime} / 100] \quad (21.7).$$

S is percentage share, * indicates summation over the e subscript it replaces and the prime indicates a revised estimate. Equation (21.3) sums the interpolated LAD mid-year populations over ethnicity, (21.4) computes the percentage shares, (21.5) carries out the extrapolation, (21.6) makes sure the shares add up to 100 and (21.7) generates the extrapolated populations.

We handle the oldest cohorts using the same equations as for the standard cohorts. Prior to these computations we must estimate the populations by single year of age over 100, which are not available at LAD scale. We use populations estimated by ONS (2015b) of the very old (including centenarians), decomposed by single years of age from 90 to 104 with 105+ as the final age group for 2002 to 2014 to decompose LAD populations aged 100+:

$$P^{i}_{g,x,e} = P^{i}_{g,100+,e} \times [P^{UK}_{g,x}/P^{UK}_{g,100+}], \text{ x=100, … , 104, 105+} \quad (21.8).$$

We assume that all persons aged 105+ are 105 and that no one is 106+.

From the sequence of estimated LAD mid-year populations by gender, age and ethnicity, we computed PAR to use with component rates described in section 21.6. The PAR are required for mid-year to mid-year intervals and are computed as averages of pairs of mid-year population estimates, as shown in Figure 21.3.

The PAR for the standard period-cohorts (and the oldest period-cohorts) are computed as averages of two successive mid-year populations, as follows:

$$PAR^{i,\{y,y+1\}}_{g,x_{pc},e} = \tfrac{1}{2} \times [P^{i,y}_{g,x,e} + P^{i,y+1}_{g,x+1,e}] \quad (21.9)$$

where the pair of times, {y, y+1}, refers to a mid-year to mid-year interval and the subscript x_{pc} refers to a period-cohort with start age x. The PARs for the fertility model that estimates births by ethnicity uses a period-age PAR.

21.6 The estimation of the ethnic components of change

The methods used to estimate the components of ethnic population change between censuses are now outlined.

Fertility estimates between CBBEs

Figure 21.4 shows the sequence of computations for fertility. The CBBE fertility rates estimates for calendar year 2001 and 2011 were based on a method developed by Norman et al. (2014) for 2001 and also used for calendar year 2011. From aggregate census tables of the population by gender, age and ethnicity were extracted infants aged 0 and mothers aged 15-44 by ethnicity to compute child-woman ratios for LADs. Births data by age of mother were used with mid-year female populations by age to compute age specific and total fertility rates (no ethnicity). Census data from the Samples of Anonymised Records and from several years of the Labour Force Survey were used to compute ethnic and age specific fertility rates for England as a whole. These proxy estimates were then combined to produce CBBE estimates (pink box in Figure 21.4) of ethnic and age specific fertility rates for LADs (Norman et al., 2014, Table 4). Fertility rates for mid-year intervals between the 2001 and 2011 CBBE estimates (light orange box in Figure 21.4) were derived through linear interpolation, adapting the population interpolation method shown in equation (21.1), allowing for the switch from CBBE calendar years to mid-year to mid-year intervals:

$$f_{x,e}^{i,\{y,y+1\}} = f_{x,e}^{i,2001} + (y - \tfrac{1}{2}) \times \frac{\left[f_{x,e}^{i,2011} - f_{x,e}^{i,2001}\right]}{10} \quad \forall\, y = 1,10 \qquad (21.10).$$

Then the interpolated rates were multiplied by PAR defined by equation 21.9 (green boxes) to yield modelled births (blue boxes). Modelled births by age and ethnicity are then adjusted to sum to ONS births data for LADs provided in the official reconciled populations and components ONS (2013).

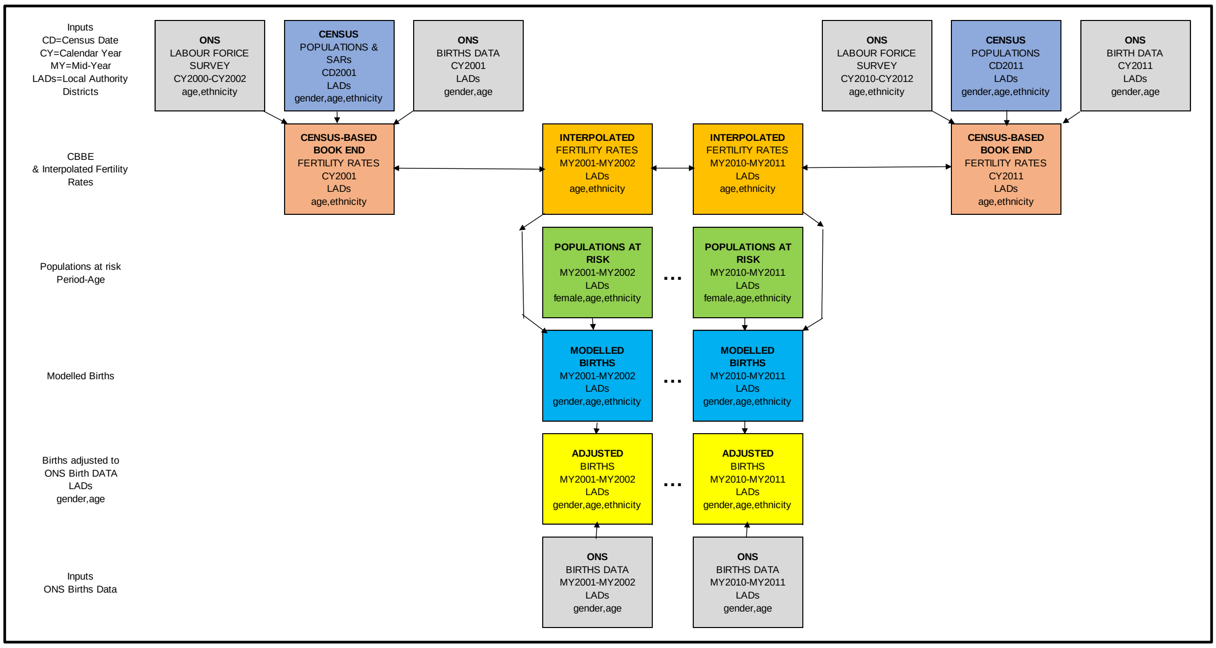


Figure 21.4 Scheme for estimating ethnic-specific fertility rates and births, mid-year 2001 to mid-year 2011

Mortality estimates between CBBEs

Figure 21.5 sets out the computations needed to estimate deaths by period-cohort and ethnicity for mid-year to mid-year intervals between CBBEs. Rees et al. (2009) estimated ethnic mortality rates and life expectancies (LEs) for LADs using two methods: the first used knowledge of long-term illness by ethnicity available in the censuses while the second used knowledge of the geographical distribution of ethnic groups. In this analysis, the geographical distribution method (GDM) was used which involved computing a national estimate of ethnic mortality by using a weighted sum of local authority mortalities, the weights being based on the geographical distribution of each ethnic group. The national estimates were used in each LAD but adjusted to match the overall mortality rates. A fully satisfactory estimate of ethnic mortality must await incorporation of ethnicity into death registration.

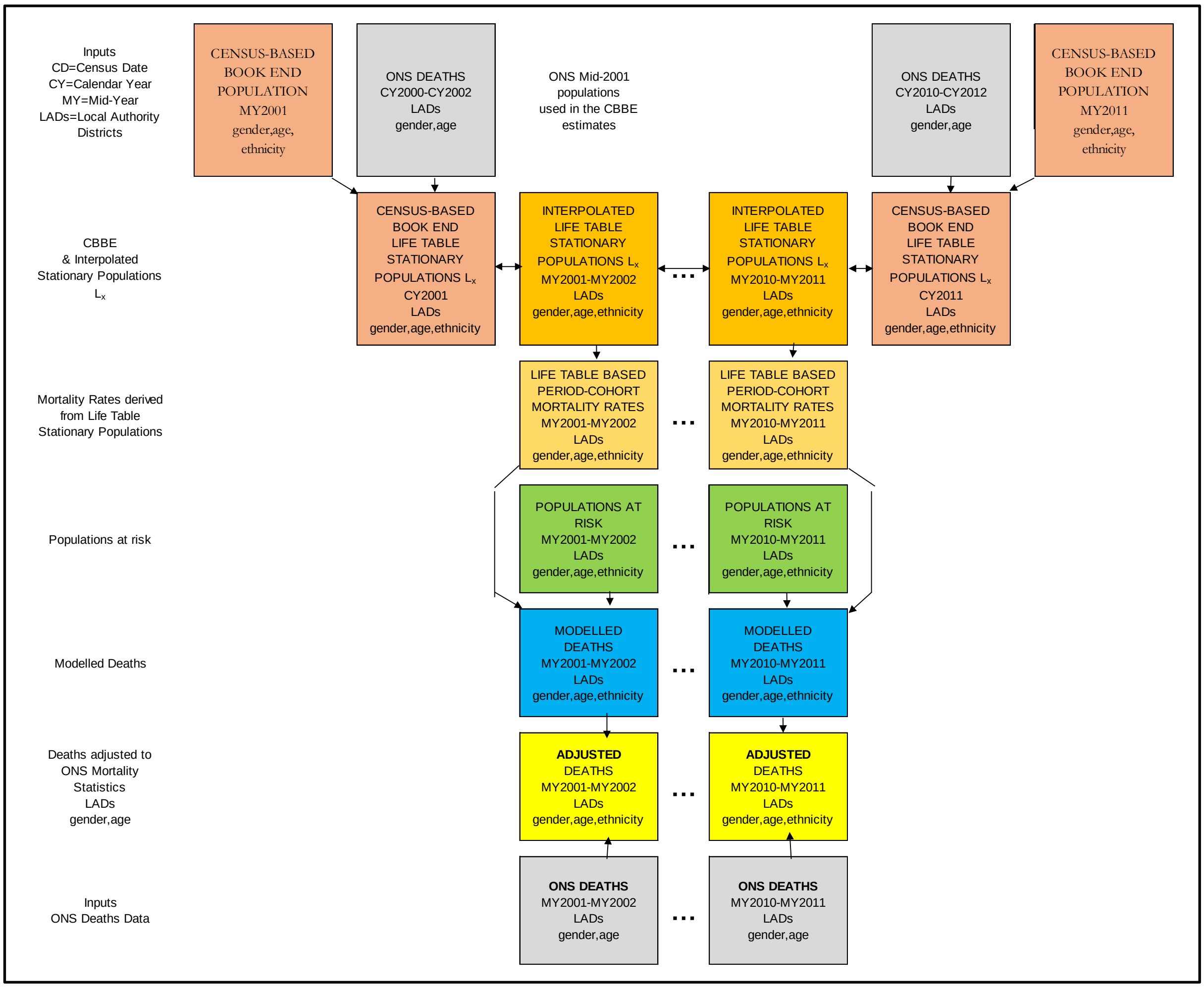


Figure 21.5 Scheme for estimating ethnic-specific mortality, mid-year 2001 to mid-year 2011

The CBBE estimates refer to the three calendar years that bracket the census, following standard ONS practice, and combine CBBE estimates for MY2001 and MY2011 of the geographical distribution of ethnic groups with the deaths data for years 2000-2002 and 2010-2012. Between CBBE estimates, mortality rates are interpolated for mid-year to mid-year intervals (Figure 21.5). The interpolated life table variable was the stationary population, L_x, for mid-year to mid-year intervals y = 0, 9 or MY2001-MY2002 to MY2010-MY2011:

$$L_{g,x,e}^{i,\{y,y+1\}} = L_{g,x,e}^{i,2001} + \left(\frac{y+0.5}{10}\right) \times \left[L_{g,x,e}^{i,2011} - L_{g,x,e}^{i,2001}\right] \qquad (21.11).$$

From these interpolated stationary populations, we computed the mortality rates needed for forming the assumptions for our ethnic group projections. The mortality rates, m, are occurrence-exposure rates for period-cohorts which are applied in the projection model to PAR averaged between successive mid-years. They are computed as follows for period-cohorts indexed by the x subscript of the mortality rate age 0-1, 1-2, … , 98-99 and age 99-100:

$$m_{g,x,e}^{i} = \left[L_{g,x,e}^{i} - L_{g,x+1,e}^{i}\right] / ½\left[L_{g,x,e}^{i} + L_{g,x+1,e}^{i}\right] \qquad (21.12).$$

For the final period-cohort, aged 100+ to 101+, the equation is modified to:

$$m_{g,100+,e}^{i} = \left[L_{g,100+,e}^{i} - L_{g,101+,e}^{i}\right] / ½\left[L_{g,100+,e}^{i} + L_{g,101+,e}^{i}\right] \qquad (21.13).$$

For the first period-cohort, the equation to uses the life table radix, l_0 , fixed at 100,000 for all population groups and an empirical factor is computed from infant mortality data for England and Wales, 2000-2014 (Rees, 2015), which estimates the fraction of a year spent alive by babies born in that year at 0.225. The mortality rates for the period-cohort linking the new-born to age 0 are estimated from:

$$m_{g,b,e}^{i} = \left[l_0 - L_{g,0,e}^{i}\right] / \left[0.225 l_0 + 0.775 L_{g,0,e}^{i}\right] \qquad (21.14).$$

Internal migration estimates between CBBEs

The scheme for estimating internal migrations between CBBEs is set out in Figure 21.6. Lomax (2013) constructed a dataset which is a complete origin by destination by age array of migration flows that harmonizes definitions between home countries and estimates missing elements such as flows between LADs in different home countries (bottom row of Figure 21.6). The technique used to create the dataset involved use of 2001 Census migration flows. This provided a complete UK origin-destination dataset for the year prior to the 2001Census as initial estimates of flows between LADs for intermediate years from 2000-2001 to 2010-2011. The census flows were adjusted using iterative proportional fitting methods (Lomax and Norman,

2016) to inter-census marginal flow data from NHS datasets recording changes of patient addresses.

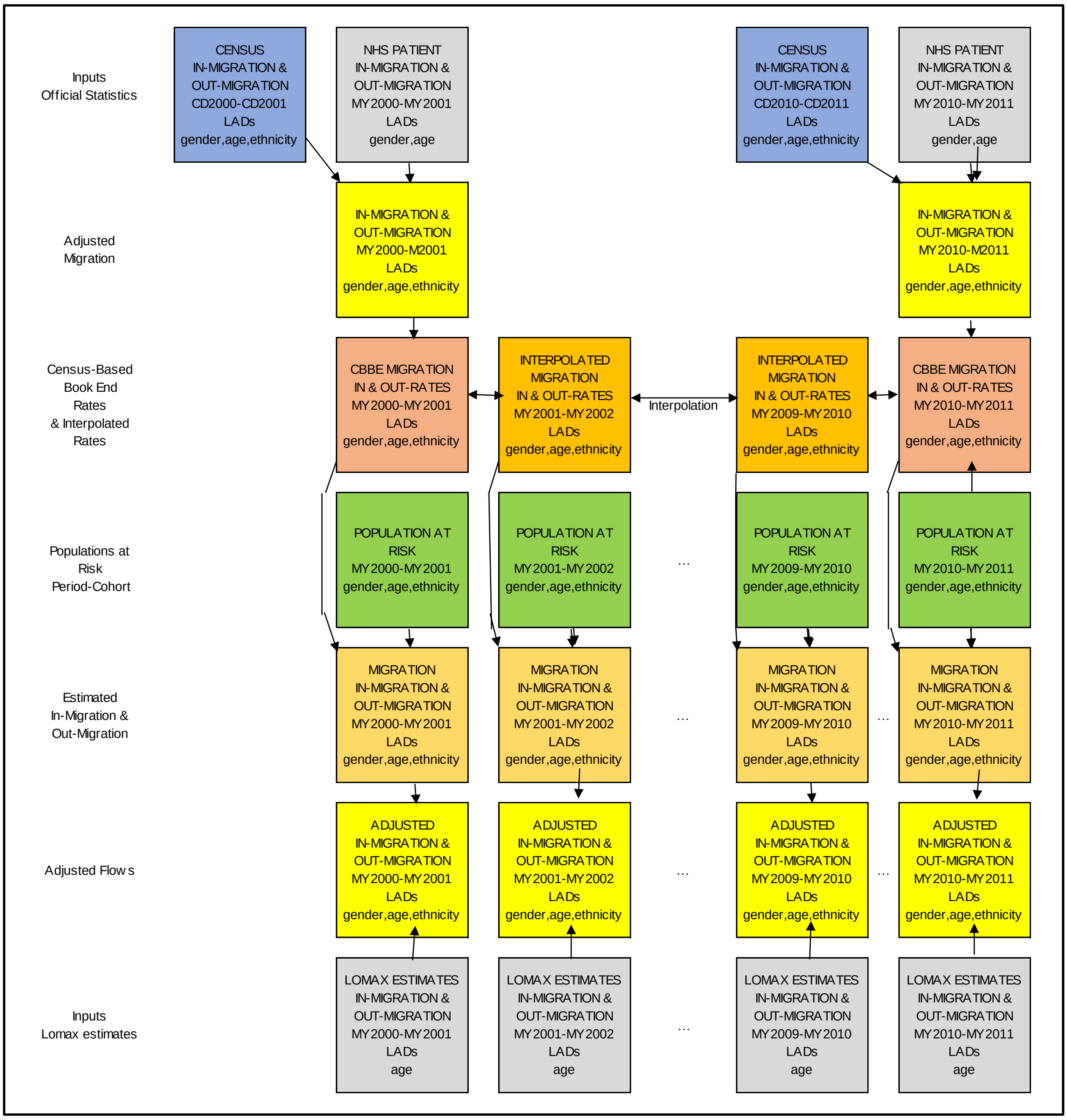


Figure 21.6 Scheme for estimating ethnic-specific internal migration, mid-year 2001 to mid-year 2011

The estimations have been extended to use census migration data for 2011 as well as 2001 and have been linked to available census cross-tabulations of LAD to LAD flows by ethnicity, making assumptions when a full classification by ethnicity could not be released because of disclosure rules (Figure 21.6, rows 1 and 2). For use in experimental projections (Rees et al., 2015a) which used the bi-regional projection model, the flow arrays were aggregated to sets of LAD inflows and outflows from/to the rest of the UK. These flows were used to estimate in- and out-migration for the CBBE intervals MY2000 to MY2001 and MY2010 to MY2011 and divided by the appropriate PAR to yield in-migration and out-migration rates. Thereafter, in- and out-migration rates for intermediate mid-year to mid-year periods were interpolated using the methods set out in section 21.5.

The interpolated rates were employed along with PAR to produce estimates of internal migration inflows and outflows (Figure 21.6, row 4) and a final adjustment was made to align the ethnic specific LAD migration inflows and outflows with estimates for all groups. The inflows and outflows in this case were not extracted from the ONS (2013) reconciled populations and components dataset but from the updated estimates made by Lomax, which dealt with cross-border flows more accurately. The final adjustment built in the known variation in the volume and directions of internal migration during the decade, associated with the economic conditions in each year – the boom years of mid-2000 to mid-2007 followed by the severe recession of mid-2007 to mid-2010, followed by a slow recovery (Lomax et al., 2013; 2014).

International migration estimates between CBBEs

Figure 21.7 presents the scheme for estimating international migration for LADs. For this component we estimate the flows of immigrants and emigrants rather than rates. The reason for estimating flows rather than rates is that in the planned projections, we have left open the form in which international migration is incorporated. A challenge in estimating international migration both for CBBE years and in between was assigning an ethnic composition to both national and international flows. To help with this we made use of the official International Passenger Survey (IPS) and Long Term International Migration (LTIM).

Census migration tables provide immigration flow data for the single years prior to the 2001 and 2011 Censuses, classified by ethnicity. No equivalent information is available for emigration, so we modelled these flows by applying ONS emigration rates for LAD populations (not distinguishing ethnicity) by the census population classified by ethnicity. ONS emigration rates are themselves a modelled estimate dependent on the LAD population, immigration in previous years and various attributes. The CBBE immigration and emigration flows were then interpolated using equations equivalent to those for the population (section 21.5).

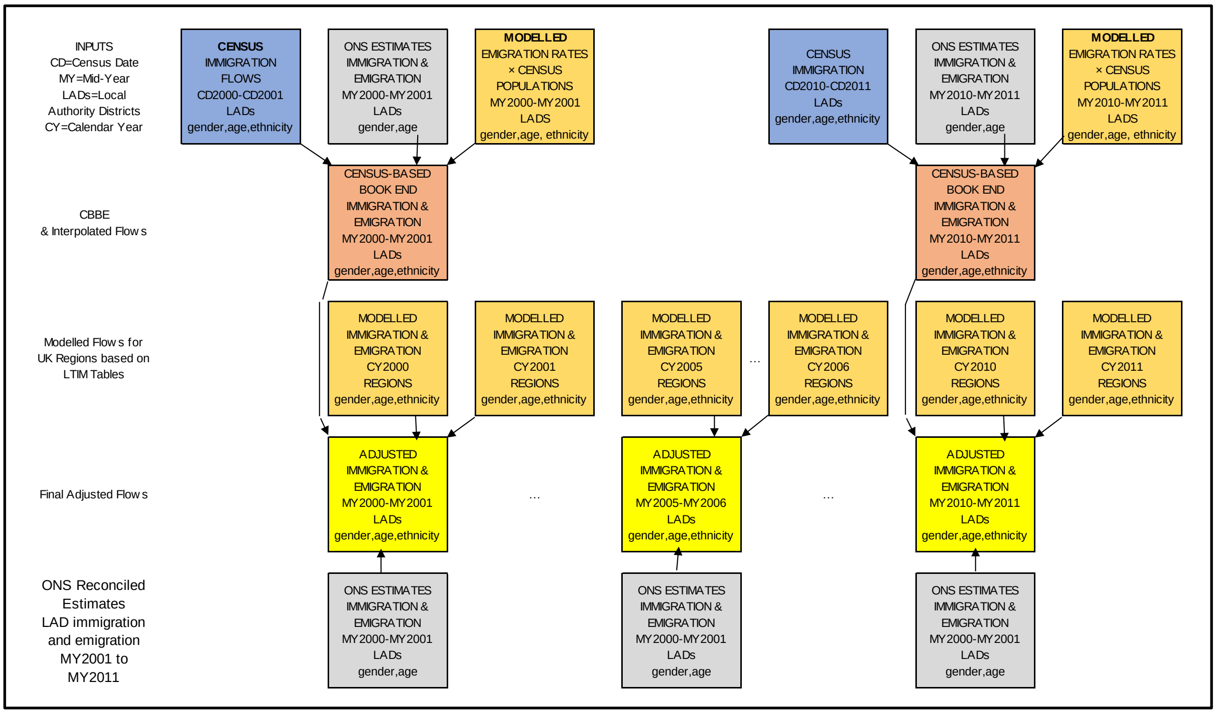


Figure 21.7 Scheme for estimating ethnic-specific international migration, mid-year 2001 to mid-year 2011

IPS and LTIM data for 2000 to 2011 point to considerable variation in the country of origin of immigrants with a major increase in immigrants from Eastern Europe occurring after eight countries joined the EU in May 2004 and rising numbers of immigrants from Southern Europe as a result of the global financial crisis starting in 2008. Using data from the IPS/LTIM data series, we estimated a time series of immigration and emigration flows for calendar years 2000 to 2014, using new definitions of country of birth groups introduced in 2014. Using 2001 and 2011 Census tables of country of birth by ethnicity we computed the probability that a person born in a country group would be a member of one the 12 ethnic groups. Multiplication of the country of birth time series by these probabilities yielded immigration and emigration totals by ethnicity for the nine regions of England and for Wales, Scotland and Northern Ireland. The LAD scale immigration and emigration flows were controlled to these regional compositions by ethnicity to produce adjusted international migration flows.

21.7 Demographic accounts for reconciling ethnic population and components

When the estimation phase of the components of change between CBBEs has been run, the next step was the reconciliation of the component information with the population changes. Demographic accounts (‘movement based’) are assembled from the populations and component estimates for each LAD, ethnicity, age and sex, mid-year to mid-year interval in order to investigate whether the components when added or subtracted from the start population yield the final population. This is equivalent to summing the inputs (start population, in-migrations, immigration) and outputs (deaths, emigrations, out-migrations and final population). The accounts can also be represented as a vector connecting start and final populations in a time interval.

When the inputs to a demographic account do not equal the outputs, then adjustments must be made to the component estimates. There are many ways in which an arithmetic balance could be achieved. We make judgements, varying by age, about how the closure error across components should be distributed, based on knowledge of the data sources and estimation methods. We assume the bulk of the closure error derives from international migration and that emigration is far more uncertain than immigration. Errors in internal migration play an important part; we assign equal weights to both in- and out-migration as the data sources and estimation methods are common to the two flows. Finally, we regard the births and deaths estimates as most reliable but judge errors in deaths estimates increase at older ages.

21.8 Discussion and conclusions

In this chapter we have described methods used to estimate ethnic population change in UK local areas. It will be possible to look at annual trends and fluctuations in the growth of minority ethnic groups and at any declines in the combined White British and Irish majority group at the geographical scale of LAD. The new estimates will be crucial for forming projection assumptions for components. In this process the availability of census datasets which reliably measure the population distribution and age-sex structure of ethnic group populations was vital. Ethnicity is not recorded on a systematic basis in any of the systems used to gather component change data (Norman et al., 2010). The registration of births and deaths does not require the identification of the ethnicity of the new-born or of the deceased. Where such information is collected, it is on a voluntary basis or through a third party. There are difficulties in determining the ethnicity of migrants. The NHS registration system is used as a proxy for migration but patients do not have to declare their ethnic identity. The IPS only asks questions about country of previous or next residence or country of birth or citizenship. The APS, despite its name, is still focussed on gathering labour market statistics and fails to provide reliable estimates of population ethnicity outside of the big cities because of the sample size is limited. These deficiencies in capturing people's ethnicity in administrative records or surveys led us to place reliance on the results of two successive censuses which both record ethnicity.

To produce population and component estimates by ethnicity for LADs we employed an age and time framework that connects people, born into the same cohort, between censuses. However, the available sources yield demographic statistics for different time points and time intervals. A first step was to define, therefore, CBBE time points for population and CBBE time intervals for components. Two CBBE time intervals were defined: the first, employed for births and deaths was based on calendar years around 2001 and 2011; the second, for internal and international migration, used mid-year to mid-year intervals prior to the censuses. Using CBBE populations and components as anchor points we were able to estimate intervening populations and components. We reviewed possible proxy variables that might indicate how much trends in ethnic populations departed from the linear over the 10 years between CBBEs but rejected all candidates. We employed linear interpolation between CBBEs. In the case of international migration we felt a linear interpolation did not fit the story of immigration waves during the decade and developed new estimates of the ethnic composition of immigration and emigration flows based on IPS/LTIM tables. Previous work had shown how important international migration was for local ethnic population change.

The methods described here were implemented through a program written in R. The equations served as the design for writing the program; the program served as a check on the consistency of the equations. Because many readers might find these equations a block to understanding rather than an aid, we have used flow charts to show how different inputs feed into the procedures used for estimating the populations and change components. We concluded our explanation of the methodology by describing the demographic accounting frameworks into which the population estimates and components fit and pointing to the need for a final round of adjustment based on intelligent adjustment when the accounts did not precisely balance.

Without two reliable UK censuses in 2001 and 2011 none of this work would have been possible. National statistical agencies, pressured by national governments seeking to limit public expenditure, are investigating how to save on the expense of censuses through use of administrative data, surveys and ‘big data’. These sources can provide valuable univariate population attribute information (Ajebon and Norman, 2015) but not cross-tabulations, especially with an ethnic group dimension. Cutting out censuses removes the most reliable knowledge about our population and prejudices the reliability of future prognoses. Censuses tell us what kind of people live where at regular time intervals and using census information we will be able to make better forecasts of how the population will change in the future.

Acknowledgements

The research reported in this chapter was supported by ESRC award ES/L013878/1 for the project Evaluation, Revision and Extension of Ethnic Population Projections – NewETHPOP and by the University of Leeds Higher Education Innovation Fund. The datasets used in the analysis were sourced from the Office for National Statistics, National Records Scotland and the Northern Ireland Statistics & Research Agency under the terms of the UK’s Open Government Licence and are Crown Copyright.

References

Ajebon, M. and Norman, P. (2015) Beyond the census: a spatial analysis of health and deprivation in England. GeoJournal, DOI:10.1007/s10708-015-9624-8.

Bryant, J.R. & Graham, P.J. (2013) Bayesian demographic accounts: Subnational population estimation using multiple data sources. Bayesian Analysis, 8(3): 591-622.

Clark, S. (2015) Modelling the impacts of demographic ageing on the demand for health care services. PhD Dissertation, School of Geography, University of Leeds. Available at: http://etheses.whiterose.ac.uk/11676/

GLA (2015) 2014 round ethnic group population projections. Greater London Authority, London Datastore. Available at: http://data.london.gov.uk/dataset/2014-round-ethnic-group-population-projections

Large, P. and Ghosh, K. (2006a) A methodology for estimating the population by ethnic group for areas within England. Population Trends, 123: 21-39.

Large, P. and Ghosh, K. (2006b) Estimates of the population by ethnic group for areas within England. Population Trends, 124: 8-25.

Lomax, N. (2013) Internal and cross-border migration in the United Kingdom: Harmonising, estimating and analysing a decade of flow data. PhD dissertation, School of Geography, University of Leeds, Leeds. Available at: http://etheses.whiterose.ac.uk/5839/

Lomax, N., Norman, P., Rees, P. and Stillwell, J. (2013). Subnational migration in the United Kingdom: Producing a consistent time series using a combination of available data and estimates. Journal of Population Research, 30(3): 265-288.

Lomax, N., Stillwell, J., Norman, P. and Rees, P. (2014). Internal migration in the United Kingdom: Analysis of an estimated inter-district time series, 2001-2011. Applied Spatial Analysis and Policy, 7(1): 25-45.

Lomax, N. and Norman, P. (2016) Estimating population attribute values in a table: ‘Get me started in’ Iterative Proportional Fitting (IPF). Professional Geographer, DOI: 10.1080/00330124.2015.1099449.

Norman, P., Rees, P., Wohland, P. and Boden, P. (2010) Ethnic group populations: the components for projection, demographic rates and trends. In Stillwell, J. and van Ham, M. (eds.) Ethnicity and Integration. Springer, Dordrecht, pp. 289-315.

ONS (2011) Population Estimates by Ethnic Group, 2002-2009. Statistical Bulletin, Office for National Statistics. Available at: http://ons.gov.uk/ons/rel/peeg/population-estimates-by-ethnic-group--experimental-/current-estimates/population-density--change-and-concentration-in-great-britain.pdf.

ONS (2013) Population Estimates for England and Wales, Mid-2002 to Mid-2010 Revised (Subnational). Statistical Bulletin. Available at: http://www.ons.gov.uk/ons/dcp171778_307723.pdf.

ONS (2015a) ONS Census Transformation Programme: Administrative Data Update. Office for National Statistics. Available at: http://www.ons.gov.uk/ons/guide-method/census/2021-census/progress-and-development/research-projects/beyond-2011-research-and-design/research-outputs/administrative-data-update.pdf.

ONS (2015b) Estimates of the Very Old (including Centenarians), England and Wales, and United Kingdom, 2002 to 2014. Statistical Bulletin, Office for National Statistics. Available at: http://www.ons.gov.uk/ons/dcp171778_418037.pdf.

ONS (2016) Annual Population Survey. NOMIS, Official Labour Market Statistics, Office for National Statistics. Available at: https://www.nomisweb.co.uk/

Rees, P., Wohland, P. and Norman, P. (2009) The estimation of mortality for ethnic groups within the United Kingdom. Social Science and Medicine, 69(11): 1592–1607.

Rees, P., Wohland, P., Norman, P. and Boden, P. (2011) A local analysis of ethnic group population trends and projections for the UK. Journal of Population Research, 28:149-183.

Rees, P., Wohland, P., Norman, P. and Boden, P. (2012) Ethnic population projections for the UK, 2001-2051. Journal of Population Research, 29: 45-89.

Rees, P., Wohland, P., Clark, S., Lomax, N. and Norman, P. (2015) Ethnic Population Projections for the UK and Local Areas, 2011-2061: New Results for the Fourth Demographic Transition. Presentation, British Society for Population Studies, Annual Conference, Leeds, 7-9 September 2015. Available at: https://www.ethpop.org/Presentations/NewETHPOP/BSPS2015/BSPS%20Ethnic%20Population%20Projections%20Phil%20Rees.pdf

Rees, P, Wohland P & Norman P (2016) The United Kingdom's multi-ethnic future: how fast is it arriving? In Lombard J, Stern E, Clarke G P (ed.) Applied Spatial Modelling and Planning, Routledge, London.

Wohland, P., Rees, P., Norman, P., Boden, P. and Jasinska, M. (2010) Ethnic population projections for the UK and local areas, 2001-2051. Working Paper 10/02, School of Geography, University of Leeds, Leeds. Available at:

http://www.geog.leeds.ac.uk/fileadmin/documents/research/csap/working_papers_new/2010-02.pdf.

Wohland, P., Rees, P., Nazroo, J. and Jagger, C. (2015) Inequalities in healthy life expectancy between ethnic groups in England and Wales in 2001. Ethnicity and Health, 20(4): 341-353