

Update 15-2017

Ethnic Group Population Projection Model Methodology

November 2017

Summary

This Update describes the operation of the ethnic group projection model used in the production of the 2016-based round of population projections.

Up to 2015 (the 2014 Round of projections) the GLA produced ethnic group population projections on the basis of ten aggregated ethnic groups. For the 2015 Round (published in November 2016) the ethnic model was extended to 17 ethnic groups. The 17 groups are the 18 ONS ethnic groups recorded in the 2011 Census, with the exception that the White Gypsy or Irish Traveller group has been aggregated into the Other White group as it is too small to model on its own (see Appendix 1 for definitions).

The ethnic model distributes the numbers of males and females for each borough from the GLA main borough model into 17 ethnic groups. It does this by generating numbers of males and females by ethnicity and then constraining them to the main borough male and female outputs.

For the 2016-based round of ethnic group projections two variants have been produced: a housing-led projection and a central trend projection. These projections differ only in the underlying borough projection to which they are constrained. The ethnic group model methodology is the same for both. Details of how the borough-level trend and housing-led projections differ can be found in their respective methodology documents¹.

At the London level, the housing-led model is constrained to the central trend variant. This means that total populations by sex and age for London are the same across the two ethnic group model variants. However, due to the way the model operates individual ethnic group totals are different between the two variants.

¹ <https://data.london.gov.uk/dataset/projections-documentation>

Background

The GLA Demography Team produce a range of annually updated non-ethnic demographic projections at both borough and ward level for the 33 local authorities in the London region. Each round of projections includes a number of variants designed to meet various user requirements, supporting the authority, its functional bodies and London boroughs in planning and policy decision making.

In addition, for the most relevant variants (usually two each year) ethnic group population projections at borough level are produced, using the demographic projections as constraints. Ethnic projections were initially produced between 1995 and 2000 by the London Research Centre (LRC), in terms of the ten ethnic groups used by ONS in 1991 Census outputs. Publication of borough ethnic projections on the basis of these ten ethnic groups has continued, in most years, after the LRC was absorbed into the GLA at its formation in 2000.

After the 2001 Census, which provided sixteen categories to choose from in the ethnicity question, one consideration was whether to maintain continuity with the 1991 Census outputs and GLA ethnic projections based on them. In addition, due to the counts of some of the sixteen categories being small it was considered necessary to use ethnic group aggregation in order to diminish the impact of statistical disclosure control applied in the 2001 census borough-level ethnic outputs published by ONS or commissioned from them. Hence ethnic model design and outputs were maintained at the ten-group level.

The 2011 Census provided eighteen groups to choose from in the ethnicity question. Whilst small counts were still present in the published and commissioned ethnicity outputs these were noticeably reduced in quantity due to higher overall ethnic populations. Furthermore, the statistical disclosure control applied by ONS to these outputs used a different methodology than for 2001 census outputs and it was considered that this would have a reduced impact on the GLA ethnic modelling process. There is also increased interest from users in having projected ethnic populations available for all of the 2011 Census ethnic groups.

From the 2015 Round onwards the ethnic model was extended to 17 ethnic groups. The 17 groups are the 18 ONS ethnic groups used in the 2011 Census, with the exception that the White Gypsy or Irish Traveller group has been aggregated into the Other White group as it is too small to model on its own (see Appendix 1 for definitions).

The ethnic model distributes the numbers of males and females by single-year-of-age for each borough from the main borough model into 17 ethnic groups. It does this by generating outputs of numbers of males and females by single-year-of-age, for each of the 17 ethnic groups, for each borough and projection year, and then constraining to the GLA main model male and female single-year-of-age figures for each borough for the given projection year. Most, but not all, ethnic inputs correspond to the borough, male, female, ethnic group, single-year-of-age structure.

Overview of Methodology

The methodology used by the ethnic model is cohort-component projection incorporating births, deaths, and migration flows in a year's population to estimate the population for the following year. These are structured by borough, ethnic group, sex, and single-year-of-age.

The base (starting) ethnic populations are 2011 Census populations by sex and single-year of age, rolled forward from Census Day to Mid-Year using ONS 2011 MYE (Mid-Year Estimate) populations, with GLA-adjusted 0 to 3-year olds, as constraints.

Ethnic age-specific fertility rates are based on delivery episodes from 2002-03 Hospital Episode Statistics and 2001 Census ethnic female populations, and are held constant over the projection period. These are used to produce births by ethnicity of mother, which are then converted to births by ethnicity of infant. This is achieved using a "transgenerational ethnicity" process involving probabilities and odds ratios of an infant's ethnicity depending on the ethnicities of the mother and father based on 2011 Census data.

The Vital Statistics Registration process in England & Wales does not collect information on ethnicity. Therefore there is no established reliable information on which to base any possible differentials in ethnic mortality rates. Although country of birth is collected in the Vital Statistics Registration process it cannot be used as a proxy for ethnicity because significant numbers of descendants of foreign-born parents are born in England & Wales. Hence the ethnic model assumes all ethnic groups have the same mortality age structures. The non-ethnic mortality structures used are those produced by the main borough model for each borough, by sex and single-year-of-age, and are specific to each projection year.

Rates and probabilities of ethnic migration are based on ethnic moves by sex and age captured in the 2011 Census, and are held constant over the projection period.

Base Ethnic Populations

Base 2011 Mid-Year Ethnic Populations by sex and single-year-of-age:

Ethnic 2011 Census Day populations were derived from 2011 Census borough-level eighteen ethnic group populations by sex and single-year of age published by ONS as 2011 Census *Commissioned Tables CT0109 - Sex by age by ethnic group (18 groups)* and *CT0122 - Age by ethnic group by sex* (both tables were commissioned for London boroughs only). The single-years-of-age in table CT0109 are from 0 up to 74, thereafter older ages are grouped as 75+, and City of London and Westminster are separate. Table CT0122 has more breakdown at the older ages, single-years-of-age from 0 to 84, then 5-year groups 85-89, 90-94, 95-99 and finally 100+, but City of London and Westminster are aggregated. Hence the two tables were used in conjunction to estimate separate older populations for City of London and Westminster by single-years-of-age from 75 up to 89 and the 90+ group. For the remaining boroughs table CT0122 was used to estimate older populations at single-years-of-age from 85 up to 89 and the 90+ group.

These were rolled forward, at ethnic sex and single-year of age level, from Census Day to Mid-Year using GLA-adjusted ONS 2011 MYE populations by sex and single-year of age as constraints. The GLA-adjusted MYE populations form the basis of all current GLA demographic projections, and incorporate the addition of 13,068 0 to 3-year olds to the Greater London ONS MYE, increasing it from 8,204,407 to 8,217,475. This GLA adjustment was first implemented in the R2013 borough non-ethnic projections in order to better reflect the pattern of births in each borough than was implied by census counts of 0 to 3-year olds. These rolled-forward populations formed the base ethnic populations for the 17-group model, with the Gypsy or Irish Traveller group being aggregated into the Other White group.

The 2011 base ethnic populations used in the new 17-group model are consistent with the 2011 base ethnic populations used in the preceding R2014 ethnic ten-group model.

Births

Ethnic Fertility Rates

Ethnic age-specific fertility rates by five-year age-band of mother used in the 17-group model were estimated from corresponding rates previously derived for the ten-group model, by allocating the rates of each aggregated ethnic group in Appendix 2 to its constituent census ethnic groups. These rates were held constant over the projection period.

To estimate ethnic fertility rates it is necessary to have available robust ethnic populations of females of child-bearing age. This information is only available from a Census, hence the timeliness of deriving ethnic fertility rates is limited in this respect. Another limitation is information on ethnic births. The Vital Statistics Registration process in England & Wales does not collect information on ethnicity and therefore ONS are unable to publish births by ethnicity based directly on Birth Registration information.

Therefore London-level ethnic age-specific fertility rates used in all preceding ten-group ethnic models were derived using deliveries by ethnicity of mother from Hospital Episode Statistics (HES) for 2002/03 as a proxy for ethnic births. These were defined in terms of the 16 2001 Census ethnic groups and grouped into the seven five-year age bands 15-19 ... 45-49 of child-bearing age of mother. Corresponding populations of females of child-bearing age were obtained from the ONS 2002 Mid-Year Estimate for London, distributed into the 16 2001 Census ethnic groups using the sex and age structures in 2001 Census Table ST101. The HES deliveries and female populations were aggregated into the ten GLA aggregated ethnic groups in Appendix 2 (the Gypsy or Irish Traveller and Arab ethnic groups were not defined separately in the 2001 Census). London-level age-specific fertility rates by five-year age band of mother for each aggregated ethnic group were then derived. This derivation of ethnic age-specific fertility rates is fully described in *DMAG Briefing 2004-24, Fertility of Ethnic Groups in London, 2002-03*.

GLA are looking into updating ethnic age-specific fertility rates for the 17-group model using HES data on deliveries by ethnicity of mother for 2011.

Births Modelling Process

Data from the 2001 and 2011 Censuses indicate that the ethnicity of a child can differ from that of its parents, particularly from its mother in the context of ethnic births. Therefore in the 2006 Round of Ethnic Projections the concept of transgenerational ethnicity of births was introduced into the modelling process to take this into account. The resulting and current births modelling process consists of three steps, the derivation of initial total (male + female) births by ethnicity of mother and constraining thereof, the application of transgenerational ethnicity to convert total births by ethnicity of mother to total births by ethnicity of infant, and the apportionment of total births by ethnicity of infant into male and female. These three steps are applied to each year of projection.

Step 1 - Initial births

Initial births (male + female) by age band and ethnicity of mother in each borough for a given projection year were derived from London-level age-specific fertility rates and the average of the female populations at the beginning and end of the projection year. This was at the 15-19 ... 45-49 five-year age-band level for each ethnic group, i.e. five-year ethnic age-specific fertility rates were applied to corresponding five-year ethnic populations of females of child-bearing age. Each number of initial births was then constrained by the borough number of total births for the corresponding projection year from the 2016-based main borough demographic model. This in effect adjusted the London-level ethnic age-specific fertility rates to borough-level ethnic age-specific fertility rates.

Using notation consistent with *DMAG Briefing 2005/21, Transgenerational Ethnicity*², this process can be described as follows:

$$B_{init}^B(x, e_1) = P_W^B(x, e_1) * f(x, e_1) \dots\dots\dots(1)$$

- where B_{init}^B = initial number of births (male + female) at borough level;
- P_W^B = number of women at borough level;
- f = age-specific fertility rate (5-year age-bands);
- x = age band of mother;
- e_1 = ethnicity of mother.

Thus $B_{init}^B(x, e_1)$ is the initial births to mothers aged x and ethnicity e_1 ; $P_W^B(x, e_1)$ is the number of women aged x and ethnicity e_1 ; and $f(x, e_1)$ is the age-specific fertility rate of women of ethnicity e_1 . Each number of constrained initial births, by age band and ethnicity of mother, $B_{cons}^B(x, e_1)$, is then given by:

$$B_{cons}^B(x, e_1) = B_{init}^B(x, e_1) * B_{DEM}^B / \sum_{e_1} [B_{init}^B(x, e_1)] \dots\dots\dots(2)$$

where B_{DEM}^B is the borough number of total births from the main borough demographic model.

Step 2 - Transgenerational ethnicity of births

The concept of transgenerational ethnicity and the technical details of its components and their derivation are described fully in *DMAG Briefing 2005/21, Transgenerational Ethnicity*. This section gives an overview of the concept and describes its application in the ethnic model.

Transgenerational ethnicity is a method of apportioning the constrained total births in Step 1 into ethnicities of the infant that takes into account the ethnicities of both the mother and father. It has two components, the probability that the infant is of a certain ethnicity given the mother's ethnicity and age and the father's ethnicity, and the likelihood of the father being of a certain ethnicity given the mother's ethnicity and age.

2011 Census Commissioned Table CT0365 relates the ethnicities of mothers, fathers and infants (children under 1 year old) in London, during the year preceding the Census. In this table counts for the Census Gypsy or Irish Traveller group have been aggregated into the Other White group by ONS to minimise disclosure control requirements. Hence the information corresponds directly with the 17 ethnic groups in the GLA 17-group ethnic model. This information was used to derive probabilities for ethnicity of the infant depending on ethnicities of the mother and father (first transgenerational ethnicity component).

Using notation consistent with *Briefing 2005/21*, these probabilities can be defined as:

$$C(e_2/x, e_1, e_3) = I(e_2/x, e_1, e_3) / \sum_{e_2} [I(e_2/x, e_1, e_3)] \dots\dots\dots(3)$$

- where C = probability;
- I = number of infants from Table CT0365 (male + female);

² <https://data.london.gov.uk/dataset/transgenerational-ethnicity>

x = age band of mother;
 e_1, e_2, e_3 = ethnicities of mother, infant and father respectively.

Thus, $C(e_2/x, e_1, e_3)$ is the probability that the infant will be of ethnicity e_2 if its mother is aged x and of ethnicity e_1 , and its father is of ethnicity e_3 .

For a given ethnicity and age band of mother there are 17 possible ethnicities of partner (father) and 17 possible ethnicities of infant, giving rise to a matrix of 289 probabilities of ethnicity of infant v ethnicity of father. Following the method of *Briefing 2005/21* in terms of age, six age bands of mother (five five-year and one 40+) were used for each of the 17 ethnicities of mother, therefore 102 matrices, each of 289 probabilities, were derived. These matrices were integrated into the ethnic model. As the fertility component of the ethnic model uses seven five-year age bands of mother (40-44 and 45+ instead of 40+) the 17 matrices corresponding to 40+ mothers were applied to both 40-44 mothers and 45+ mothers in the model, so that the 17-group ethnic model incorporates 119 matrices in total.

The second transgenerational ethnicity component, the likelihood of a father being of a certain ethnicity, can be considered to depend on the “pool” or number of available males of that ethnicity and their age, in relation to ethnicity and age of the mother. Following the method of *Briefing 2005/21*, the 2011 Census outcomes of father’s ethnicity in relation to mother’s ethnicity and age in Table CT0365 were used in conjunction with Census population data from Table DC2101EW indicating potential fathers to derive relative risks for the ethnicity of the father. For each five-year age-band of mother the pool of potential fathers was taken as a “corresponding” 10-year age-band male population, broken down by ethnicity. So for mothers aged 15-19 the pool of potential fathers was from the 15-24 male age band, and similarly through to mothers aged 45-49 for whom the pool of potential fathers was from the 45-54 male age band.

Hence,

$$R_{x, e_1}(e_3) = \frac{M(e_3/x, e_1)}{Q(e_3/x, e_1)} = \frac{K_{x, e_1}(X, e_3) / K_{x, e_1}(X, A)}{P_{M, x, e_1}(X, e_3) / P_{M, x, e_1}(X, A)} \dots\dots\dots(4)$$

- where R = relative risk of ethnicity of father;
- M = probability of ethnicity of father;
- Q = relative pool of potential fathers by ethnicity;
- K = number of fathers of known ethnicity (from Census Commissioned Table CT0365);
- P_M = number of males (from Census Table DC2101EW);
- X = wider age band of males associated with age band of mother x ;
- A denotes “all” (i.e. fathers, or males).

$R_{x, e_1}(e_3)$ is the relative risk of males of ethnicity e_3 being fathers of children borne by women of age x and ethnicity e_1 .

$M(e_3/x, e_1)$ is the probability that the father of a child borne by a woman aged x and of ethnicity e_1 will be of ethnicity e_3 .

$Q(e_3/x, e_1)$ is the relative pool of potential fathers of ethnicity e_3 within the overall pool of males of age X associated with mothers aged x and of ethnicity e_1 .

$K_{x, e_1}(X, e_3)$ and $K_{x, e_1}(X, A)$ are respectively the number of fathers of ethnicity e_3 and total number of fathers of known ethnicity in the group of fathers of age X associated with mothers aged x and of ethnicity e_1 .

$P_{M, x, e_1}(X, e_3)$ and $P_{M, x, e_1}(X, A)$ are respectively the number of males of ethnicity e_3 and total number of males in the pool of males of age X associated with mothers aged x and of ethnicity e_1 .

However, because the information in Commissioned Table CT0365 was not perfect and the ethnicity of the father was not always known, it was necessary to adjust the relative risks to reflect the complete number of fathers (i.e. infants) in CT0365. This also had to take account of cases where all fathers within a breakdown of infant’s ethnicity were not known, i.e. there was no known father’s ethnicity to scale with. The process

used for adjustment was to apply the derived values of $R_{x, e1}(e_3)$, $M(e_3/x, e_1)$, and $C(e_2/x, e_1, e_3)$ to the total number of infants, within each ethnicity and age band of mother, to predict ethnicity of all infants, in doing so minimising the differences between predicted and actual infant sub-totals of each ethnicity. This was done using the Solver optimising function in Microsoft Excel. The process is described in more detail in *Briefing 2005/21*. Resulting adjusted relative risks, denoted as $RR_{x, e1}(e_3)$, were derived for each ethnicity of father in relation to ethnicity and age band of mother (for six age bands of mother, five five-year and one 40+, as for the parameter $C(e_2/x, e_1, e_3)$ described above). Again, as the fertility component of the ethnic model uses seven five-year age bands of mother (40-44 and 45+ instead of 40+), the adjusted relative risks corresponding to 40+ mothers were applied to both 40-44 mothers and 45+ mothers in the model. In the model the adjusted relative risks were formatted in 119 17 x 1 matrices corresponding to the 119 matrices for the parameter $C(e_2/x, e_1, e_3)$, each matrix relating to an ethnicity and age band of mother, and consisting of 17 ethnicities of father v one adjusted relative risk.

The parameters $RR_{x, e1}(e_3)$, $M(e_3/x, e_1)$, and $C(e_2/x, e_1, e_3)$ were derived using London-level data, the level at which Commissioned Table CT0365 was commissioned. It was impractical to commission at borough level because counts would have been much smaller and a much greater amount of data would have been subject to disclosure control, resulting in low robustness in any modelling process. Hence the objective of the transgenerational ethnicity methodology was to derive the parameters at London-level and apply them at borough level. The London-level values of parameter $C(e_2/x, e_1, e_3)$ can be regarded as stable, both through time and for different areas, but the parameter $M(e_3/x, e_1)$ is likely to vary with the characteristics of the male population. Equation (4) can be re-arranged to express the probability of father's ethnicity at borough level, M^B , in terms of borough-level relative pools of fathers Q^B or numbers of males $P^B_{M, x, e1}$, to reflect differences in characteristics of the male population at borough level when applied in the ethnic model:

$$M^B(e_3/x, e_1) = RR_{x, e1}(e_3) * Q^B(e_3/x, e_1) = RR_{x, e1}(e_3) * P^B_{M, x, e1}(X, e_3) / P^B_{M, x, e1}(X, A) \dots\dots(5)$$

The "transgenerational probability", or the probability that a birth to a woman aged x and of ethnicity e_1 , and whose partner is of ethnicity e_3 , will itself be ascribed to ethnicity e_2 , denoted by $T^B(x, e_1, e_3, e_2)$ at borough level (otherwise $T(x, e_1, e_3, e_2)$ at London level) can then be derived from:

$$T^B(x, e_1, e_3, e_2) = C(e_2/x, e_1, e_3) * M^B(e_3/x, e_1)$$

ie

$$T^B(x, e_1, e_3, e_2) = C(e_2/x, e_1, e_3) * RR_{x, e1}(e_3) * P^B_{M, x, e1}(X, e_3) / P^B_{M, x, e1}(X, A) \dots\dots\dots(6)$$

The transgenerational probability from Equation (6) can then be applied to the constrained initial births derived by Equation (2) to distribute the latter into their own ethnicities, at borough level:

$$B^B(x, e_1, e_3, e_2) = B^B_{cons}(x, e_1) * T^B(x, e_1, e_3, e_2)$$

ie

$$B^B(x, e_1, e_3, e_2) = B^B_{cons}(x, e_1) * C(e_2/x, e_1, e_3) * RR_{x, e1}(e_3) * P^B_{M, x, e1}(X, e_3) / P^B_{M, x, e1}(X, A) \dots\dots\dots(7)$$

In the ethnic model, Equation (7) was used to incorporate 119 matrices, each of 289 outcomes of ethnicity of births, i.e. each of 17 ethnicities of births v 17 ethnicities of father, corresponding to the 119 matrices of the parameter $C(e_2/x, e1, e3)$ and of the adjusted relative risks $RR_{x, e1}(e_3)$, and relating to the 119 groupings of ethnicity and age band of mother. Within each matrix the value of each outcome was the number of births of a given ethnicity, and hence the total number of births of any given ethnicity arising from mothers of any given ethnicity and age band was the sum of the 17 outcomes of the given ethnicity of births, and similarly the sum of all 289 outcomes was the total number of births associated with the mother of the said ethnicity and age band. It should be noted that within each matrix the values of the majority of the 289 outcomes were 0 or very small.

So the sum of 17 matrix outcomes for a given ethnicity of birth was described by:

$$B^B(x, e_1, A, e_2) = \sum_{e_3} [B^B(x, e_1, e_3, e_2)] \dots\dots\dots(8)$$

where A denotes all ethnicities of father.

And the sum of 289 matrix outcomes for all ethnicities of births was described by:

$$B^B(x, e_1, A, A) = \sum_{e_3} \sum_{e_2} [B^B(x, e_1, e_3, e_2)] \dots\dots\dots(9)$$

where the additional A denotes all ethnicities of births e_2 , and $B^B(x, e_1, A, A)$ originated from $B^B_{cons}(x, e_1)$ in Equation (2), the initial constrained births to a woman of ethnicity e_1 and age band x .

It should be noted at this stage that the total number of births of a given ethnicity in the borough (the sum of 17 matrix outcomes for a given ethnicity of births in all 119 matrices) would be described by:

$$B^B(A, A, A, e_2) = \sum_x \sum_{e_1} \sum_{e_3} [B^B(x, e_1, e_3, e_2)] \dots\dots\dots(10)$$

And the sum of all births in the borough (the sum of all 289 matrix outcomes in all 119 matrices) would be described by:

$$B^B(A, A, A, A) = \sum_x \sum_{e_1} \sum_{e_3} \sum_{e_2} [B^B(x, e_1, e_3, e_2)] \dots\dots\dots(11)$$

However, because the probabilities for father’s ethnicity, parameter $M(e_3/ x, e_1)$ in Equation (4), were derived using London-level relative ethnic pools of males, unless borough-level relative ethnic pools of males happened to be the same then at borough level a matrix sum of all 289 birth outcomes associated with mothers of age band x and ethnicity e_1 , $B^B(x, e_1, A, A)$, derived by Equation (9) would not be equal to the original initial constrained births to these mothers, $B^B_{cons}(x, e_1)$, derived by Equation (2). This result was inevitable due to the impracticability of commissioning Census data at borough level instead of at London level, as stated previously, and therefore only having London-level probabilities of father’s ethnicity at one’s disposal.

It was considered that the outcomes within a given matrix gave the best available distribution of transgenerational ethnicities of births, and that it was acceptable to constrain each component within a matrix (derived by Equation (7)) to return to the initial constrained total births for that matrix (derived as $B^B_{cons}(x, e_1)$ in Equation (2)). Hence, individual constrained matrix outcomes were described by:

$$B^B_{cons}(x, e_1, e_3, e_2) = B^B(x, e_1, e_3, e_2) * B^B_{cons}(x, e_1) / B^B(x, e_1, A, A) \dots\dots\dots(12)$$

The individual constrained matrix outcomes $B^B_{cons}(x, e_1, e_3, e_2)$ can be summed as in Equations (8), (9), (10) and (11) to give corresponding sub-totals or totals.

Hence, applying the summation criteria of Equation (10) to Equation (12), the (final) transgenerational total number of births of a given ethnicity in the borough would be described by:

$$B^B_{cons}(A, A, A, e_2) = \sum_x \sum_{e_1} \sum_{e_3} [B^B_{cons}(x, e_1, e_3, e_2)] \dots\dots\dots(13)$$

And, applying the summation criteria of Equation (11) to Equation (12), the (final) transgenerational total number of births in the borough would be described by:

$$B^B_{cons}(A, A, A, A) = \sum_x \sum_{e_1} \sum_{e_3} \sum_{e_2} [B^B_{cons}(x, e_1, e_3, e_2)] \dots\dots\dots(14)$$

The process of implementing transgenerational ethnicity described above was repeated for each projection year.

In summary, as stated before, the parameter $C(e_2/x, e_1, e_3)$, describing probability of ethnicity of birth depending on age and ethnicity of mother and ethnicity of father, was derived from Census data and was considered to be stable through time and for different areas. Hence, the 119 matrices of $C(e_2/x, e_1, e_3)$ that were applied to each borough and each projection year were the same. The parameter $M(e_3/ x, e_1)$, the probability of ethnicity of father given the ethnicity of infant and ethnicity and age of mother, was also derived from Census data at London level, and was used to derive London-level adjusted relative risks of males of a given ethnicity being fathers of children borne by women of given ethnicity and age. The

London-level adjusted relative risks were applied to each borough and in each projection year. However, the probability of ethnicity of father was applied in borough-level terms as the parameter $M^B(e_3/x, e_1)$, in conjunction with the London-level adjusted relative risks, as in Equation (5). In other words, the other inputs to $M^B(e_3/x, e_1)$ were borough-specific and projection year-specific male populations, and this had the effect of converting the London-level probabilities of ethnicity of father into borough-level ones, and of updating them each projection year to reflect changes in the borough male population. This process resulted in a new distribution of ethnicity of births for each ethnicity and age grouping of mother, which was then constrained to give the corresponding number of total births for each ethnicity and age grouping of mother at the start of the process.

Step 3 – Apportionment of ethnic total births into male and female

Up to 2015 (the 2014 Round of projections) the GLA ethnic group population projection methodology used ethnic-specific birth sex ratios derived from the ratio of census ethnic male and female 0-year olds for Greater London, which implied that birth sex ratio differed by ethnicity.

However, recent analysis published by the Department of Health, *Birth Ratios in England and Wales - A report on gender ratios at birth in England and Wales, May 2014*, investigated birth sex ratios in terms of country of birth of mother and in terms of ethnicity of infant. No country of birth of mother or ethnicity of infant was found to have a ratio that was statistically significantly different from the range that might be expected to occur naturally for births overall.

Therefore the GLA 17-group ethnic model uses an assumed naturally-occurring ratio, with no ethnic or regional variation. This has been implemented for the 2015 Round onwards and brings the GLA ethnic projection model into line with the GLA main borough model. The naturally-occurring ratio is the ratio of UK births in the ONS 2014-based NPP (National Populations Projections) fertility assumptions, which is 105 live male births per 100 live female births. This birth sex ratio is used for each borough and each projection year.

Hence the transgenerational total number of births of a given ethnicity in a borough, derived by Equation (13), was apportioned into ethnic male births by:

$$\text{Male [} B_{\text{cons}}^B(A, A, A, e_2) \text{]} = B_{\text{cons}}^B(A, A, A, e_2) * \frac{105}{205} \dots\dots\dots(15)$$

And the transgenerational total number of births of a given ethnicity in a borough, derived by Equation (13), was apportioned into ethnic female births by:

$$\text{Female [} B_{\text{cons}}^B(A, A, A, e_2) \text{]} = B_{\text{cons}}^B(A, A, A, e_2) * \frac{100}{205} \dots\dots\dots(16)$$

Therefore approximately 51.22 per cent of borough total births in each ethnic group are apportioned as males and 48.78 per cent as females.

Deaths/Survivors

The Vital Statistics Registration process in England & Wales does not collect information on ethnicity and therefore ONS are unable to publish deaths by ethnicity based directly on Death Registration information. Hence there is no reliable information on possible variation in ethnic mortality and the GLA 17-group ethnic model assumes there is none.

Therefore specific survival rates by borough, sex, single-year-of-age, and projection year are used to derive numbers of deaths and survivors. These survival rates are produced by the GLA main borough model. The resulting ethnic deaths are constrained to corresponding total borough deaths for each projection year, also produced by the main borough model.

Deaths and Survivors for age 0 were derived from births:

$$\text{Deaths}_{\text{borough, sex, year, } 0\text{ya}} = \text{Births}_{\text{borough, sex, year}} * (1 - \text{Survival Rate}_{\text{borough, sex, year, } 0\text{ya}})$$

$$\text{Survivors}_{\text{borough, sex, year, } 0\text{ya}} = \text{Births}_{\text{borough, sex, year}} * \text{Survival Rate}_{\text{borough, sex, year, } 0\text{ya}}$$

Deaths and Survivors for ages 1 to 89 were derived from preceding age population in preceding year:

$$\text{Deaths}_{\text{borough, sex, year, } \text{sya}1 \dots 89} = \text{Population}_{\text{borough, sex, year-1, } \text{sya-1}} * (1 - \text{Survival Rate}_{\text{borough, sex, year, } \text{sya}})$$

$$\text{Survivors}_{\text{borough, sex, year, } \text{sya}1 \dots 89} = \text{Population}_{\text{borough, sex, year-1, } \text{sya-1}} * \text{Survival Rate}_{\text{borough, sex, year, } \text{sya}}$$

Deaths and Survivors for age 90+ were derived from preceding age (89) population and age 90+ population in preceding year:

$$\text{Deaths}_{\text{borough, sex, year, } \text{sya } 90+} = (\text{Population}_{\text{borough, sex, year-1, } \text{sya-1}} + \text{Population}_{\text{borough, sex, year-1, } \text{sya } 90+}) * (1 - \text{Survival Rate}_{\text{borough, sex, year, } \text{sya } 90+})$$

$$\text{Survivors}_{\text{borough, sex, year, } \text{sya } 90+} = (\text{Population}_{\text{borough, sex, year-1, } \text{sya-1}} + \text{Population}_{\text{borough, sex, year-1, } \text{sya } 90+}) * \text{Survival Rate}_{\text{borough, sex, year, } \text{sya } 90+}$$

Migration In - From UK

Information on in-migration moves, from anywhere in the UK (including other London boroughs) to each London borough, within the year up to 2011 Census Day is available from 2011 Census *Commissioned Table CT0404 - Origin and destination of migrants in to London by age by sex by ethnic groups*. The published age breakdown of these moves is restricted to five-year age bands (0, 1-4, 5-9 ... 85-89, 90+) for reasons of disclosure control, but all eighteen census ethnic groups are present. For the 17-group ethnic model it was necessary to aggregate counts for the Gypsy or Irish Traveller group into the Other White group.

The moves were used to derive ethnic in-migration rates From UK. However, where the total number of male or female moves in any ethnic group in any borough was less than 250, London-level ethnic group age structures were applied to re-distribute the moves by age on a more robust basis. This smoothing was necessary because in these cases the numbers of moves within the older age bands of some ethnic groups were very low or zero.

Using the in-moves (re-distributed where applicable), separate male and female in-migration rates From UK for each five-year age band, ethnic group, and borough were derived. For a given borough the rate denominators were the total male moves and total female moves in all ethnic groups in the borough. For implementation in the ethnic model single-year-of-age rates were estimated by allocating the rate from the corresponding age band and dividing by the number of single-years-of-age in the age band.

The male and female ethnic Census-based single-year-of-age rates were kept constant for each projection year. They were multiplied by the respective total male and total female borough From UK in-moves output by the main borough demographic model for each projection year, to give the number of From UK in-migrants for each projection year, for each single-year-of-age of males, females, ethnic group, and borough. These ethnic in-migrants were therefore automatically constrained to the main borough model.

Migration In - From Overseas

Information on in-migration moves, from Overseas to each London borough, within the year up to 2011 Census Day is available from 2011 Census *Commissioned Table CT0404 - Origin and destination of migrants in to London by age by sex by ethnic groups*. As for the From UK moves in *Commissioned Table CT0404* the published age breakdown of these moves is restricted to five-year age bands (0, 1-4, 5-9 ... 85-89, 90+) for reasons of disclosure control, but all eighteen census ethnic groups are present. For the 17-group ethnic model it was necessary to aggregate counts for the Gypsy or Irish Traveller group into the Other White group.

The moves were used to derive ethnic in-migration rates From Overseas. However, where the total number of male or female moves in any ethnic group in any borough was less than 250, London-level ethnic group age structures were applied to re-distribute the moves by age on a more robust basis. This smoothing was necessary because in these cases the numbers of moves within the older age bands of some ethnic groups were very low or zero.

Using the in-moves (re-distributed where applicable), separate male and female in-migration rates From Overseas for each five-year age band, ethnic group, and borough were derived. For a given borough the rate denominators were the total male moves and total female moves in all ethnic groups in the borough. For implementation in the ethnic model single-year-of-age rates were estimated by allocating the rate from the corresponding age band and dividing by the number of single-years-of-age in the age band.

The borough male and female ethnic Census-based single-year-of-age rates were kept constant for each projection year. They were multiplied by the respective total male and total female borough From Overseas in-moves output by the main borough demographic model for each projection year, to give the number of From Overseas in-migrants for each projection year, for each single-year-of-age of males, females, ethnic group, and borough. These ethnic in-migrants were therefore automatically constrained to the main borough model.

Migration Out - To UK

Information on out-migration moves, from each London borough to the Rest of England & Wales excluding other London boroughs, within the year up to 2011 Census Day is available from 2011 Census *Commissioned Table CT0405 - Origin and destination of migrants out of London by age by sex by ethnic groups*. Information on moves between London boroughs within the year up to 2011 Census Day is available from 2011 Census *Commissioned Table CT0404 - Origin and destination of migrants in to London by age by sex by ethnic groups* (as mentioned in the section on Migration In - From UK). The moves from these two sources were added to arrive at out-migration moves from each London borough to the Rest of England & Wales (anywhere in England & Wales including the rest of London). The published age breakdown of these moves in both sources is restricted to five-year age bands (0, 1-4, 5-9 ... 85-89, 90+) for reasons of disclosure control, but all eighteen census ethnic groups are present. For the 17-group ethnic model it was necessary to aggregate counts for the Gypsy or Irish Traveller group into the Other White group.

Borough Census Day ethnic populations were derived using 2011 Census *Commissioned Tables CT0109 - Sex by age by ethnic group (18 groups)* and *CT0122 - Age by ethnic group by sex*, but with GLA adjustment of 0 to 3-year olds by scaling-back from the GLA-adjusted 2011 MYE described in the Base Ethnic Populations section (i.e. these were not the same as the base ethnic 2011 Mid-Year GLA-adjusted populations). Both Commissioned Tables are largely by single-year-of-age and all eighteen census ethnic groups are present. For the 17-group ethnic model it was necessary to aggregate counts for the Gypsy or Irish Traveller group into the Other White group. The ages were aggregated to the five-year age bands corresponding to the out-migration moves from *Commissioned Tables CT0404 and CT0405*.

Probabilities of out-migration to the Rest of England & Wales were derived for each age band (and hence each single-year-of-age within it) by dividing the out-moves to Rest of England & Wales for each sex, ethnic group, and borough by the corresponding population. However, for most of the ethnic groups the numbers of out-moves and/or corresponding populations in older age bands was very low or zero.

The process was repeated for Inner London and Outer London. The Inner London and Outer London ethnic probability structures, resulting from sufficiently large numbers of out-moves and corresponding populations, were much more robust and less erratic than at borough level, but differences in ethnic structures between Inner and Outer London were evident.

Therefore the structures of Inner London and Outer London ethnic out-moves were used to smooth out the out-move structures in each corresponding borough for each sex and ethnic group except White British. Final ethnic probabilities of out-migration to Rest of England & Wales were derived by dividing the smoothed To Rest of England & Wales out-moves for each sex, ethnic group, and borough by the corresponding population. In the case of the White British group the numbers of out-moves and corresponding populations in each age band were always sufficiently large to result in acceptable probability structures and no smoothing was required. The resulting borough male and female ethnic Census-based single-year-of-age probabilities were kept constant for each projection year.

For the White British ethnic group the unconstrained numbers of out-migrants To Rest of England & Wales were derived by multiplying the unsmoothed single-year-of-age probabilities by the corresponding number of survivors for the projection year.

For all other ethnic groups the unconstrained numbers of out-migrants To Rest of England & Wales were derived by multiplying the smoothed single-year-of-age probabilities by the corresponding number of survivors for the projection year.

For migration in-flows the Census is able to capture moves originating anywhere in the world. Hence implementation of in-flows in the ethnic model in terms of From UK and From Overseas, as described in the preceding two sections, reflects the complete picture of in-migration. For migration out-flows the Census is only able to capture destinations within England & Wales. Hence the out-migration probabilities described

in this section are To Rest of England & Wales, and therefore exclude Scotland and Northern Ireland. However, as moves from England & Wales (particularly from London) to Scotland and Northern Ireland are comparatively low the ethnic To Rest of England & Wales out-migration probabilities can be used as an estimate of ethnic out-migration probabilities to anywhere in the United Kingdom (i.e. including Scotland and Northern Ireland). This allowed the derived ethnic numbers of out-migrants as described in this section to be constrained to the main borough model, which outputs borough total out-migration moves To UK by sex and projection year.

For males the constraining factor was the total borough male out-moves To UK from the main model divided by the total borough male ethnic out-migrants To UK from the ethnic model. Similarly, for females the constraining factor was the total borough female out-moves To UK from the main model divided by the total borough ethnic female out-migrants To UK from the ethnic model. The constrained number (for each single-year-of-age) was then derived by multiplying the unconstrained number (derived in terms of destination as Rest of England & Wales) by the constraining factor (representing destination as UK), for males and females respectively. The resulting constrained outflows were estimates of ethnic out-migration To UK.

Migration Out - To Overseas

As mentioned in the preceding section, for migration out-flows the England & Wales Census is only able to capture destinations within England & Wales. There is therefore no census data available to describe ethnic out-migration to Overseas. The ethnic model therefore used the ethnic out-migration probabilities To UK (derived in terms of destination as Rest of England & Wales), derived as described in the preceding section (smoothed for all ethnic groups except White British), as an estimate of ethnic out-migration probabilities To Overseas, but resulting ethnic numbers of out-migrants were constrained to borough total out-migration moves To Overseas output by the main borough model. As for Migration Out - To UK, the borough male and female ethnic Census-based single-year-of-age probabilities were kept constant for each projection year.

For the White British ethnic group the unconstrained numbers of out-migrants To Overseas were derived by multiplying the unsmoothed single-year-of-age probabilities by the corresponding number of survivors for the projection year.

For all other ethnic groups the unconstrained numbers of out-migrants To Overseas were derived by multiplying the smoothed single-year-of-age probabilities by the corresponding number of survivors for the projection year.

For males the constraining factor was the total borough male out-moves To Overseas from the main model divided by the total borough male ethnic out-migrants To Overseas from the ethnic model. Similarly, for females the constraining factor was the total borough female out-moves To Overseas from the main model divided by the total borough ethnic female out-migrants To Overseas from the ethnic model. The constrained number (for each single-year-of-age) was then derived by multiplying the unconstrained number by the constraining factor (representing destination as Overseas), for males and females respectively. The resulting constrained outflows were estimates of ethnic out-migration To Overseas.

Unconstrained Population

For each projection year, borough, sex, ethnic group, and single-year-of-age the unconstrained population was the net summation of the demographic components derived above, i.e.:

$$\begin{aligned}
 \text{Unconstrained population} &= \text{Survivors (surviving births in case of 0-year olds, then ages 1 to 90+)} \\
 &+ \text{constrained in-migrants From UK} \\
 &+ \text{constrained in-migrants From Overseas} \\
 &- \text{constrained out-migrants To UK} \\
 &- \text{constrained out-migrants To Overseas}
 \end{aligned}$$

Constrained Population

The final step in the model was to constrain each borough total male and each borough total female single-year-of-age unconstrained population (Unconstrained P) for all ethnic groups to the corresponding single-year-of-age population figures produced by the main borough demographic model (Main Model P), by direct re-distribution.

For each projection year, borough, sex, ethnic group (eg), and single-year-of-age (sya) the constrained population (Constrained P) is given by:

$$\text{Constrained } P_{\text{borough, sex, eg, sya, year}} = \text{Unconstrained } P_{\text{borough, sex, eg, sya, year}} * \frac{\text{Main Model } P_{\text{borough, sex, sya, year}}}{\sum (\text{Unconstrained } P_{\text{borough, sex, eg, sya, year}})}$$

Appendix 1: GLA projection 17 ethnic groups and how they relate to the 2011 Census ethnic groups

ONS 2011 Census Ethnic Group (18)	ONS 2001 Census Ethnic Category (16)	GLA ethnic model	GLA ethnic model	GLA White or BAME ¹
		seventeen-group	ten-group	
		GLA Ethnic Group	GLA Aggregated Ethnic Group (AEG) ²	
White: English/Welsh/Scottish/Northern Irish/British White: Irish White: Gypsy or Irish Traveller ⁴ White: Other White	White: British White: Irish	White British White Irish	White	White
		Other White		
	White: Other White			
Black/African/Caribbean/Black British: Caribbean	Black or Black British: Caribbean	Black Caribbean	Black Caribbean	BAME
Black/African/Caribbean/Black British: African	Black or Black British: African	Black African	Black African	
Black/African/Caribbean/Black British: Other Black	Black or Black British: Other Black	Other Black	Black Other	
Mixed/multiple ethnic group: White and Black Caribbean	Mixed: White & Black Caribbean	White & Black Caribbean		
Mixed/multiple ethnic group: White and Black African	Mixed: White & Black African	White & Black African		
Asian/Asian British: Indian	Asian or Asian British: Indian	Indian	Indian	
Asian/Asian British: Pakistani	Asian or Asian British: Pakistani	Pakistani	Pakistani	
Asian/Asian British: Bangladeshi	Asian or Asian British: Bangladeshi	Bangladeshi	Bangladeshi	
Asian/Asian British: Chinese	Chinese or Other: Chinese	Chinese	Chinese	
Mixed/multiple ethnic group: White and Asian	Mixed: White & Asian	White & Asian	Other Asian ³	
Asian/Asian British: Other Asian	Asian or Asian British: Other Asian	Other Asian ³		
Mixed/multiple ethnic group: Other Mixed	Mixed: Other Mixed	Other Mixed	Other	
Other ethnic group: Arab ⁴		Arab		
Other ethnic group: Any other ethnic group	Chinese or Other: Other	Other Ethnic Group		

¹ BAME (Black, Asian and Minority Ethnic) denotes a grouping of all ethnic groups except the white groups.

² Approximate to the ten Ethnic Groups used in the 1991 Census, using best fit comparison described in DMAG Briefing 2003/9, 2001 Census Key Statistics: Ethnicity, religion and country of birth.

³ Although the same name for the Other Asian group is used in both GLA models the definitions are different.

⁴ These two 2011 Census ethnic groups (Gypsy or Irish Traveller and Arab) were not defined separately in the 2001 Census.

Two additional ethnic groups were introduced in the 2011 Census for England & Wales, “White: Gypsy or Irish Traveller”, and “Other ethnic group: Arab”. To date there is no detailed Census data or ONS studies thereof available that establish linkage between the 2011 groups and higher-level ethnic group aggregations used in relation to data from previous Censuses. To allow comparison with previous GLA ethnic projections and permit use of the GLA 10-group ethnic model the two new groups were allocated to the White and Other GLA Aggregated Ethnic Groups respectively.

Note the different terminology used by ONS in the last two censuses for the main White group: “White: British” in the 2001 Census and “White: English/Welsh/Scottish/Northern Irish/British” in the 2011 Census. It has been assumed that this is the same ethnic group reflecting people who consider themselves of White background relating to the United Kingdom of Great Britain and Northern Ireland (this excludes White people from the Republic of Ireland, who are defined as a separate group, White Irish, in both Censuses). Technically this 2011 Census group would be summarised as “White UK”, but is referred to as “White British” for consistency and continuity.

For more information please contact Ed Klodawski, GLA Intelligence
Greater London Authority, City Hall, The Queen's Walk, More London, London SE1 2AA
Tel: 020 7983 4694 e-mail: edmund.klodawski@london.gov.uk