GLAINTELLIGENCE

Update

GLA Trend-based Projection Methodology

2018-based demographic projections

November 2019

Introduction

The GLA Demography Team produce a range of annually updated population projections at local authority, ward and MSOA level for the 33 London boroughs. Each round of projections includes a number of variants designed to meet various user requirements, but in general the variants form two groups:

- Trend projections – those based purely on trends in fertility, mortality and migration;
- Housing-led projections – those which incorporate a forecast housing development trajectory.

This Update is concerned with the development of the GLA trend projection model and the production of the 2018-based trend-based population projections at local authority level. This produces outputs for all local authorities in England and Wales as well as national outputs for Norther Ireland and Scotland. All model outputs are available through the London Datastore[1].

Note: For clarity this Update refers to mid-year periods by reference to the end year. Therefore, the year mid-2001 to mid-2002 is referenced as mid-2002. Similarly, a longer period such as that between mid-2001 and mid-2014 is referenced as mid-2002 to mid-2014.

Development the GLA cohort component model

The GLA's models and assumptions continue to evolve as new data is released and each new round of projections supersedes earlier rounds.

The primary differences between this round and the last are:

- Changes to the input population estimates and components backseries (the GLA make adjustments to official ONS mid-year estimates and components of change)
- The incorporation of trend data from the 2018-based NPP replacing data from the 2016-based NPP
- The incorporation of an additional year's mid-year estimate and domestic migration data
- Local authorities in Wales are now modelled individually rather than treating Wales as a single geographic unit

[1] https://data.london.gov.uk/demography/

The 2018-based model incorporates the following data:

- 2018 Mid-year Estimate data on births, deaths, population and international migration (adjusted by the GLA)
- Single-year-of-age by sex LA-to-LA flows for the period mid-2002 to mid-2018
- The mortality and fertility trajectories from the 2018 NPP
- Age specific mortality and fertility curves from the 2016 SNPP
- Underlying data from the DCLG 2014-based Household Projections
- Underlying data from the ONS 2016-based Household Projections

Overview of methodology

These projections are produced using a cohort component projection model. Projections are produced from the starting point of an adjusted ONS Mid-Year Estimate for 2018[2].

Each subsequent year's population is generated by the same process, taking the previous year's projected population as the start point. For mid-year to mid-year periods when the total numbers of births, deaths and net migrants are known, the results may be better described as base period estimates.

The cycle of events that takes an initial local authority population and generates a projection of the subsequent year's population is described below and illustrated in the flowchart (Figure 1).

1. The cycle begins with the initial local authority populations by single-year-of-age (0 to 90+) and sex. For the first year, this is the base population, for subsequent years this is the projected population at the end of the previous cycle.

2. The starting population is aged-on and survived to the end of the year by application of age-specific mortality rates.

3. Births are calculated by applying age-specific fertility rates to the female population (aged 15-49). As births occur throughout the projection year they are calculated using a combination of the starting and the aged-on and survived female populations at the end of the year.

4. Survival rates are applied to births to project the number that will reach 'age 0' at the end of the projection year.

5. International out-migration is calculated by applying age and sex specific rates to the aged-on population and subtracting the result from the survived population.

6. Numbers of in-migrants from overseas are projected from the historic record of international migrants and a constant age and sex distribution of the totals.

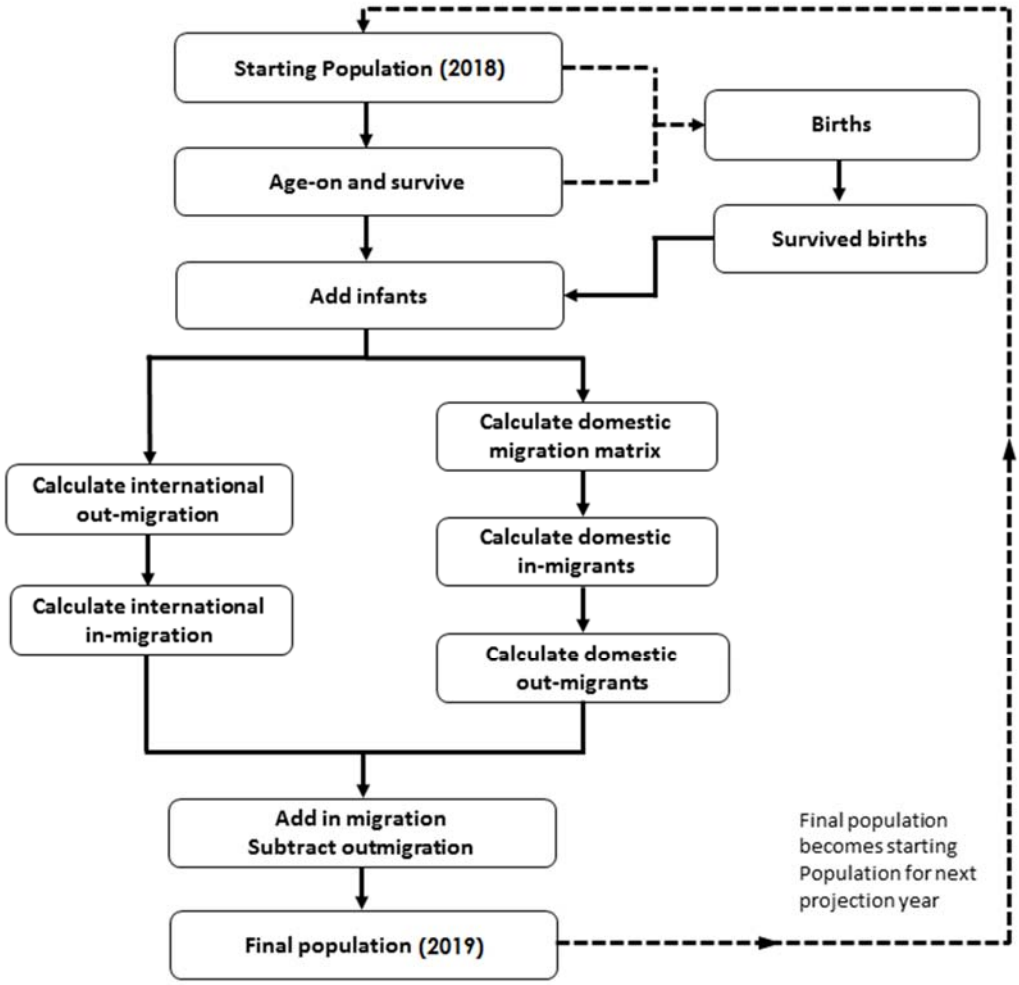
7. A domestic migration matrix is calculated by applying age and sex specific out-migration probabilities to the population. The matrix includes flows (by age and sex) between all local authorities in England and Wales as well as Northern Ireland and Scotland. Local authority-level in and out migration are calculated by summing the inflows and outflows for each authority. As with international out-migration the rates are applied to the aged-on population.

[2] See section on base population below

8. The final population for the projection year is fed back into step 1 as the initial population for the next projection year.

The model outputs estimated and projected population by single year of age and sex from 2011 to 2050. Additional reporting outputs are also produced, including births, deaths and gross migration flows.

Figure 1: Flow Chart of the Projection cycle



GLA Demography, 2019

Projection variants

Three different projection scenarios were modelled, primarily to reflect uncertainty in future migration patterns. These are labelled as the central, short-term and long-term migration scenarios, respectively. Migration flows are the most variable and difficult to project component of population change and the very large scale of flows into and out of London makes the projections especially sensitive to the assumptions used.

The short-term variant assumes that recent migration patterns will persist for the duration of the projection period. While projections based on this approach are suitable for use in the near-term, the GLA has argued

that a projection based only on recent patterns are not a suitable basis for long-term planning. When projecting further ahead it is generally better to base assumptions on longer historical trends, preferably spanning a full, or multiple, economic cycles.

For use strategic planning applications, the GLA produces outputs based on longer periods of past migration data. Incorporating a longer period of past migration patterns, reduces the influence of any individual event and yields more stable results between successive projection rounds when compared to those produced using short-term trends only.

The bases for the trends used the 2018-based scenarios are as follows:

- The short-term migration scenario bases the migration patterns on estimates for the **five-year period** mid-2014 to mid-2018.
- The central migration scenario bases the migration patterns on estimates for the **10-year period** mid-2009 to mid-2018.
- The long-term migration scenario bases the migration patterns on estimates for the **15-year period** mid-2004 to mid-2018.

The projections are otherwise the same in terms of methodologies and assumptions regarding fertility and mortality.

Internal Migration series

The ONS Internal Migration series makes available estimated local authority to local authority flows by single-year-of-age and sex for the period mid-2002 to mid-2018. The model takes advantage of this series to project flows based on 329 areas (all LAs in England and Wales plus Northern Ireland and Scotland) by single-year-of-age and sex.

The mid-2017 and mid-2018 flows estimates incorporate new input data sources and a revised estimation methodology. The Patient Demographic Service (PDS) administrative dataset has been incorporated in order to replace the NHS central Register (NHSCR) which was discontinued in February 2016.

In addition, improvements have been made to the assumptions relating to the behaviour of higher education leavers. ONS have acknowledged that the standard method for capturing internal migration moves – monitoring health registrations – works well for some groups and less well for others. In particular, young adults can be slow to register with the health service following a move and so the previous internal migration methodology was prone to understating the magnitude and distribution of post-university flows. As such, starting from the 2017 flow estimates, a new Higher Education Leavers Methodology (HELM) was implement to better account for these graduate flows[3].

The impact of the changes in 2017 is to cause something of a break in the series. The new methodology and data sources result in an increase in the number of moves overall in the UK. The impact for London of the changes varies significantly from borough to borough. Overall the difference between the old and new

[3] Full details can be found on the ONS website:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/methodologyguideformid2015ukpopulationestimatesenglandandwalesjune2016#internal-migration

methods for 2017 is an increase in inflow to London of 21,785 and an increase in outflow of 18,652. The impact on net flows is a less significant increase of 3,133.

Base Population

A series of population estimates prior to the 2018 starting population is required by the model in order to generate the rates and probabilities used to project forwards. The GLA has made a number of adjustments to the ONS mid-year estimates (MYE) back series for the 2018-based projections. These adjustments were necessary to account for systematic issues with the series as well as undercounts of young children in the 2011 census.

In particular, the GLA identified problems with the official estimates of population and migration of children for London local authorities. Analysis of the official estimates alongside GP registration data, school pupil data and births data indicated that individual cohorts of children in many boroughs were becoming inflated over time. The GLA has identified that this issue primarily results from deficiencies in the estimated migration flows in the MYE backseries.

To mitigate the impact that these issues have on the projections, the GLA has adjusted the mid-year estimates of population and international migration for children aged 0-14 over the period 2011-2018. The changes were based on a multi-stage modelling process, that sought to identify a timeseries of past population more consistent with observed trends in administrative data sources. A consistent series of international migration flows were then created based on these updated population estimates and the standard birth, death and domestic migration components.

This adjusted mid-year estimate and components series constitutes a significant change in the base data inputs to the model. These adjustments ensure that there is greater confidence in the 2018 starting population, both in terms of magnitude and age structure, and that rates derived in the model, and based on the backseries, are more robust.

Impact of changes to base population

The impact of the changes on the base population and the projections can be seen in Figure 2. In 2018, the final year of input data, the difference between the unadjusted ONS estimate and the adjusted GLA estimate for London is 96,500. The impact over the projection period is to lower the trajectory so that the projection based on the GLA adjusted series is 214,000 lower in 2041. While this is a significantly lower population, because the direct adjustment is limited to the under-15 population the impact on the projected number of households in London is relatively small. The difference between the household projections (using the DCLG household model to convert population to households) is 20,400 households in 2041 (816 per year over the 25-year period 2016-2041).

Figure 2: Difference between ONS and GLA base data

London Population, 2018 Central Trend projection

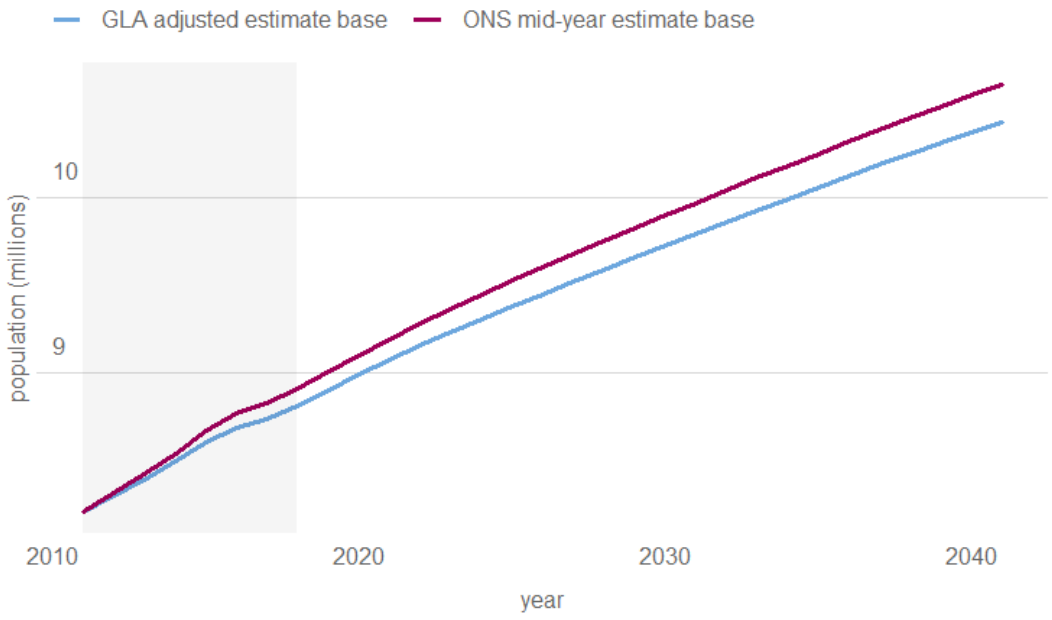


Chart: GLA City Intelligence
Source: London Datastore, ONS mid-year estimates

Births

The 2018-based projections have an updated methodology for calculating the initial age specific fertility rates (ASFR) used in the model. This has been made possible by the publication, for the first time in the 2016-based SNPP, of the underlying SNPP ASFRs.

The model takes the ASFRs for each local authority from the SNPP for the year 2017 (the SNPP's first projection year) and smooths them by fitting double-peaked Hadwiger mixture curves (see Appendix A). These ASFRs are applied to backseries populations to obtain births estimates which can be compared to recorded births. The ratio of estimated to actual births is calculated in each year and then the last five years' ratios are averaged. The average ratio is applied to the smoothed 2018 curve to produce a 2018 ASFR curve for each local authority.

Assumed fertility rates beyond 2018 follow age-specific fertility trends taken from the ONS 2018-based National Population Projections (NPP).

Detailed births Methodology

For each local authority:

1) A set of initial fertility rates for 2019 are calculated by smoothing the SNPP ASFRs for 2017 using a function which fits double-peaked Hadwiger mixture curves (see Appendix A for full details). This process modifies the rates to include ages 45-49.

2) The smoothed rates are applied to estimates of the female population to produce estimates of the total number of births for each year over the period 2014 to 2018.

3) The estimates are compared to the actual births data taken from the ONS mid-year estimates, and a local authority-level scaling factor is derived for each of the five past years.

4) The five scaling factors are averaged to give a mean scaling factor for the local authority.

5) The mean scaling factor is applied to the smoothed 2017 ASFRs to produce a set of ASFRs for 2019.

6) For years beyond 2019 the ASFRs are adjusted by applying the rate of change by single-year-of-age from the ONS 2018-based NPP principal fertility assumptions.

7) Projected births are arrived at by applying the rates in each year to the projected female population. Births assigned a sex based on the ratio of 105 males to 100 females.

Note: As births occur throughout the projection year the 'at risk' population is calculated using a combination of the starting and the aged-on and survived female populations at the end of the year. This is the case both in the calculation and application of fertility rates.

Deaths

Mortality rates are calculated using a similar method to fertility rates. In this case, the raw (unsmoothed) age-specific mortality rates (ASMR) for 2017 are taken from the outputs of the 2016-based SNPP. These are applied to mid-year populations for 2014 to 2018 and the resulting total deaths compared to recorded deaths. Scaling factors for each year are derived and the mean of these is applied to the 2017 curve to produce the 2019 ASMRs. Rates beyond 2019 follow the age-specific mortality trends taken from the 2018-based NPP.

Detailed deaths methodology

For each local authority:

1) A set of initial age-specific mortality rates for 2017 are taken from the SNPP outputs for 2017.

2) The rates are applied to estimates of the population to produce estimates of the total number of deaths for each year over the period 2014 to 2018.

3) The estimates are compared to the actual deaths data taken from the ONS mid-year estimates and a local authority-level scaling factor is derived for each year.

4) The scaling factors are trended forward using linear regression to give a calculated scaling factor for each local authority.

5) This scaling factor is applied to the 2017 ASMRs from the SNPP to produce a set of local authority ASMRs for 2019.

6) For years beyond 2019 the ASMRs are adjusted by applying the rate of change by single-year-of-age from the ONS 2018-based NPP principal mortality assumptions.

7) Projected deaths are arrived at by applying the mortality rates in each year to the projected population.

International Migration

International out-migration is calculated by applying age-specific out-migration probabilities to the population. These probabilities are the average of previous years' out-migration rate estimates:

- The short-term model variant uses five years of data (mid-2014 to mid-2018)
- The central variant uses ten years of data (mid-2009 to mid-2018)
- The long-term variant uses 15 years of data (mid-2004 to mid-2018).

International in-migration is calculated by taking an average of inflows by age and sex in previous years and holding that inflow constant for the duration of the projection.

Detailed methodology
For each local authority:

1) Base out-migration probabilities are calculated using international outmigration and population estimates from adjusted Mid-Year Estimates for the defined period (5, 10 or 15 years).

2) An average of the rates for the defined period is taken.

3) For the projection years the age and sex specific rates are applied to the population.

4) Total inflows from overseas by age and sex are taken from the adjusted Mid-Year Estimates for the defined period (5, 10 or 15 years).

5) An average of the inflow is taken.

6) For the projection years the resulting population is added to the projection.

Domestic Migration

Domestic migration is calculated using a multi-region model to project the flows between 323 areas:

- 33 London boroughs
- 293 local authorities in England
- 22 local authorities in Wales
- Northern Ireland and Scotland

Domestic migration probabilities by age and sex are calculated from ONS internal migration data and adjusted mid-year population estimates:

- The short-term model variant uses five years of data (mid-2014 to mid-2018)
- The central variant uses ten years of data (mid-2009 to mid-2018)
- The long-term variant uses 15 years (mid-2004 to mid-2018).

Detailed methodology
1) Base age and sex specific out-migration probabilities are calculated for each area pairing. These probabilities use domestic out-migration estimates from the adjusted Internal Migration series and population estimates from the adjusted Mid-Year Estimates for the defined period (5, 10 or 15 years).

For each year this creates a 4-dimensional domestic migration matrix comprising 22,295,000 rates – 350 areas x 350 areas x 91 ages x 2 sexes.

2) An average of each the rates for the defined period is taken (5, 10 or 15 years).

3) For the projection years, the age and sex specific rates are applied to the population to calculate flows between each pair of areas.

4) For each area age and gender domestic in-migration and domestic out-migration figures are calculated by summing from the individual area to area flows.

5) In-migration totals are added to the population and out-migration totals are subtracted.

Population Outputs

In addition to the projected population by single year of age (sya) and sex, the model outputs include:

- Total Births
- Births by mother's age
- ASFRs
- Total deaths (by sya/sex)
- Domestic & Internal Gross and Net flows (by sya/sex)
- Borough and regional aggregations

The release of the 2018-based projection includes projections and associated data for all local authorities in England and Wales as well as for Northern Ireland and Scotland.

Households

The GLA model produces projected households as a standard output. Households are arrived at by applying the DCLG's 2014-based household projections methodology[4] and the 2016-based ONS-based household projections methodology[5] to the projected population.

[4] The DCLG 2014 household projection methodology document can be viewed here: https://www.gov.uk/government/statistics/2014-based-household-projections-methodology

[5] The ONS 2016 household projection methodology document can be viewed here: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/datasets/householdprojectionsforengland

Appendix A: Smoothing age-specific fertility rates

Modelled age-specific fertility rates (ASFR) in the GLA model are smoothed to reduce noise. This was implemented in response to the recommendations of the review undertaken by the ESRC Centre for Population Change at the University of Southampton. This final report from this review is available on the London Datastore[6].

Base age-specific fertility rates (ASFR) are taken from the 2016-based ONS subnational population projections (SNPP). The 2016 SNPP is the first set of subnational projections to include ASFRs in its outputs.

For the 2017-based projection, a smoothing process has been applied to local authority-level SNPP ASFR curves in advance of their incorporation into the model. This process is intended to reduce random fluctuations in the data. The approach used is to use a function to fit double-peaked Hadwiger mixture curves (Chandola *et al*. 1999)[7] to the age-specific fertility rates.

The double-peak Hadwiger curve is given by the expression:

$$f(x) = am\left(\frac{b_1}{c_1}\right)\left(\frac{c_1}{x}\right)^{3/2}\exp\left\{-b_1^2\left(\frac{c_1}{x}+\frac{x}{c_1}-2\right)\right\} + (1-m)\left(\frac{b_2}{c_2}\right)\left(\frac{c_2}{x}\right)^{3/2}\exp\left\{-b_2^2\left(\frac{c_2}{x}+\frac{x}{c_2}-2\right)\right\}$$

Where:
x is age of the mother at birth

(x) is the fertility rate at age x f

m is a mixture parameter that determines the relative sizes of the two component distributions

a, b_1, c_1, b_2, c_2 are other model parameters related to total fertility and the level and trend of the mean ages of fertility in the two component distributions.

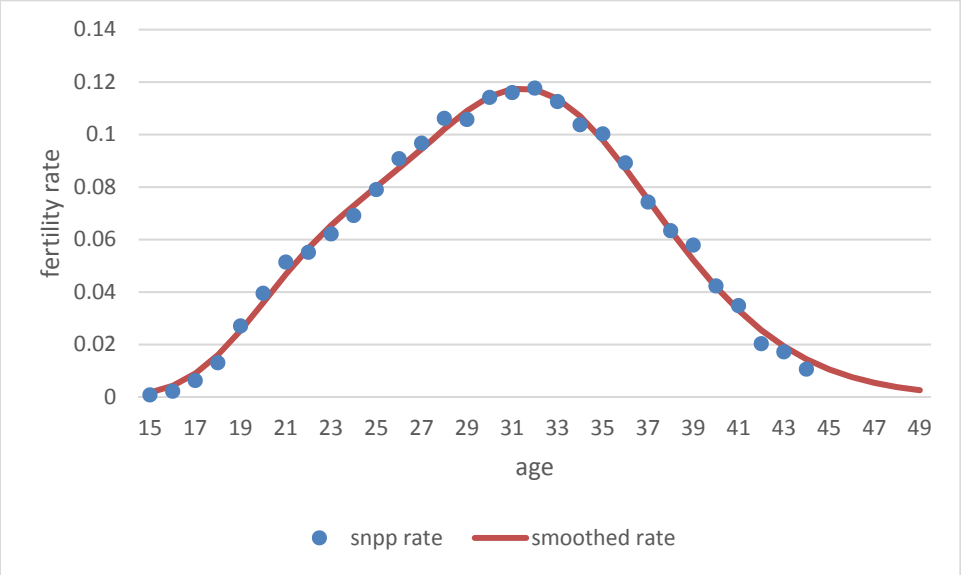
The input data to the function is a set of ASFR for the year 2017 for each local authority for ages 15 to 44 (the range for which ONS outputs ASFR data from the 2016-based SNPP). The output is a set of smoothed rates for ages 15 to 49, with ages 45 to 49 calculated by extrapolation of the fitted curve (see Figure 4).

A Levenberg-Marquardt Nonlinear Least-Squares algorithm is used to fit a curve to each set of data, with starting points chosen based on previously fitted curves from 2011, and convergence tests as described here: https://cran.r-project.org/web/packages/minpack.lm/minpack.lm.pdf using the package defaults for convergence conditions. The parameters of the fitted curve are then used to calculate new rates for each age.

If convergence does not happen within 200 iterations, a grid search method is used to run the Levenberg-Marquardt Nonlinear Least-Squares algorithm with a range of starting values in order to find the best fit. If no fit is found then the data for that local authority is left unchanged.

[6] https://data.london.gov.uk/dataset/projection-methodology-independent-review
[7] Chandola, T., Coleman, D.A., Horns, R.W. (1999) Recent European fertility patterns: fitting curves to 'distorted' distributions. Population Studies, 53, 3:317-329.

Figure 2: Raw and smoothed fertility rates for Brent, 2017



GLAINTELLIGENCE

For more information please contact: GLA Intelligence

Wil Tonkiss,
Greater London Authority,
City Hall,
The Queen's Walk,
More London,
London SE1 2AA

e-mail: demography@london.gov.uk

MAYOR OF LONDON