Update 04-2016

Estimating births using GP registration data

July 2016

Key Findings

- The GLA has developed and tested a simple methodology for using GP registration data to predict subsequent birth estimates;
- The method provides an indication of annual births almost a year ahead of official estimates provided by ONS. Timely birth estimates are important for school place planners;
- Estimates of births for the year to mid-2015 were produced for London local authorities using this method. 91% of these estimates lay within 2% of the subsequent official estimates;
- Comparison of the accuracy of the approach compares very favourably with existing methodologies used by the GLA and the ONS in their published projections;
- New estimates of births for the year to mid-2016 have been produced based on recently obtained GP registration data. These indicate London-wide births of 129,900, a rise of 1,400 compared to the previous 12 month period;
- The estimates suggest rises in births of over 200 in each of the east London boroughs of Barking & Dagenham, Newham, Havering, and Hackney.

Introduction

In January 2016, the GLA Demography Team published a technical note[1] on the use of GP registration data to provide more timely estimates of live births in London. The note presented: a methodology for producing estimates; analysis of the expected accuracy of estimates produced in this way; and a set of estimates for births for the year to mid-2015. Since this note was published, ONS has released the 2015 mid-year estimates[2], which include estimated births over the preceding year by local authority of residence.

This Update reviews the accuracy of the birth estimates produced in the earlier technical note and compares the accuracy of the approach with established projection methods used by the GLA and ONS.

New estimates of births for the year to mid-2016 have been produced, based on recently received GP registration data.

Details of the methodology and data sources used are not reproduced here, but can be found in the technical note.

---

[1] http://data.london.gov.uk/dataset/estimating-births-using-gp-registration-data
[2] https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/latest

Accuracy of birth predictions for the year to mid-2015

Following the release of the 2015 mid-year estimates, the predictions of births produced from analysis of GP registration data were compared with official estimates for the period[3]. For London as a whole, the predicted number of births was 128,054, compared with actuals of 128,520. The prediction was 466 lower than the actuals, a difference of 0.36%. Differences for individual local authorities were larger and are set out in Table 1, overleaf.

The original analysis included estimates of expected distribution of the accuracy of estimates produced using the methodology. The distributions of anticipated and actual accuracy of the results are plotted in Figures 1 and 2, respectively. Accuracy of the results proved to be better than expected, with a smaller distribution of variances around the actual estimates. Analysis had suggested that 84% of predictions would fall within 2% of the actual estimates. For mid-2015, 91% of predictions fell within this margin of error.

To put the accuracy of the approach into context, the results can be compared with other sources of predicted births for the period. Here we will consider the results of the latest ONS and GLA projections, both of which include projected numbers of births at local authority level for 2015 and both of which were based on actual birth estimates up to and including those for the year to mid-2014. The methodology documents issued by ONS[4] and the GLA[5] explain how births are projected in each case.

It should be noted that these are *projections* rather than *predictions* and that expectations of accuracy are not necessarily the same. However, the relative accuracy of the results compared to the GP register-based predictions is important as it informs the GLA's decision about whether to incorporate these data into its demographic projection models. If the accuracy of these predictions is significantly better than the equivalent projected values, then there is a case to include the data as part of the estimates series from which the model projects forward.

The accuracy of the three different approaches for each borough is shown in Table 2. Figures 3 to 5 illustrate the distributions of the accuracy of the results for each source. There are various statistical methods for determining whether one set of predictions is better than another. Looking at the overall agreement at London-level does not provide an adequate benchmark. By this measure, the standard GLA projection approach produced a closer result than the GP-register based method. However, this single figure hides larger borough-level errors, which tend to cancel when aggregated to London-level. A better approach might be to compare the mean squared error (MSE) for each set of predictions. This is calculated as follows:

$$Mean\ Squared\ Error = \frac{1}{n}\sum_{i=1}^{n}(P_i - A_i)^2$$

Where n is the number of predictions, and P and A are the predicted and actual values, respectively.

A value of zero implies perfect accuracy and larger values indicate greater overall error in the predictions made. Figure 6 shows the calculated MSE values for the 32 London authorities (Hackney and City of London combined) for each set of predictions. This indicates that the predictions made using counts from the GP register are more accurate than the results of both the ONS and GLA's standard projection methodologies.

---

[3] Note: results for City of London have been combined with those for Hackney in all subsequent analysis of accuracy.

[4] http://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/qmis/subnationalpopulationprojectionsqmi

[5] https://files.datapress.com/london/dataset/2015-round-population-projections/2016-07-04T14:15:19/GLA%20trend%20projection%20methodology.pdf

Table 1: Comparison of actual and predicted mid-2015 birth estimates

Local authority	Predicted 2015 mid-year birth estimate	Actual 2015 mid-year birth estimate	Absolute error	Percentage error
Barking & Dagenham	3,685	3,697	-12	-0.3%
Barnet	5,251	5,222	29	0.6%
Bexley	3,122	3,093	29	0.9%
Brent	5,092	5,157	-65	-1.3%
Bromley	4,036	4,098	-62	-1.5%
Camden	2,754	2,735	19	0.7%
Croydon	5,606	5,705	-99	-1.7%
Ealing	5,430	5,389	41	0.8%
Enfield	4,973	4,895	78	1.6%
Greenwich	4,567	4,547	20	0.4%
Hammersmith & Fulham	2,394	2,371	23	1.0%
Haringey	4,176	4,112	64	1.6%
Harrow	3,443	3,509	-66	-1.9%
Havering	3,081	3,138	-57	-1.8%
Hillingdon	4,467	4,482	-15	-0.3%
Hounslow	4,349	4,339	10	0.2%
Islington	2,859	2,908	-49	-1.7%
Kensington & Chelsea	1,823	1,823	0	0.0%
Kingston upon Thames	2,323	2,314	9	0.4%
Lambeth	4,562	4,549	13	0.3%
Lewisham	4,848	4,763	85	1.8%
Merton	3,307	3,330	-23	-0.7%
Newham	5,860	5,998	-138	-2.3%
Redbridge	4,770	4,801	-31	-0.6%
Richmond upon Thames	2,610	2,621	-11	-0.4%
Southwark	4,620	4,641	-21	-0.5%
Sutton	2,788	2,822	-34	-1.2%
Tower Hamlets	4,529	4,587	-58	-1.3%
Waltham Forest	4,610	4,670	-60	-1.3%
Wandsworth	5,008	5,057	-49	-1.0%
Westminster	2,663	2,685	-22	-0.8%
City and Hackney	4,420	4,462	-42	-0.9%
Greater London	128,054	128,520	-466	-0.4%

Source: ONS mid-year estimates; ONS 2014-based subnational projections; GLA 2015 round short term trend projection

Figure 1: Expected distribution of accuracy of birth predictions using GP registration data
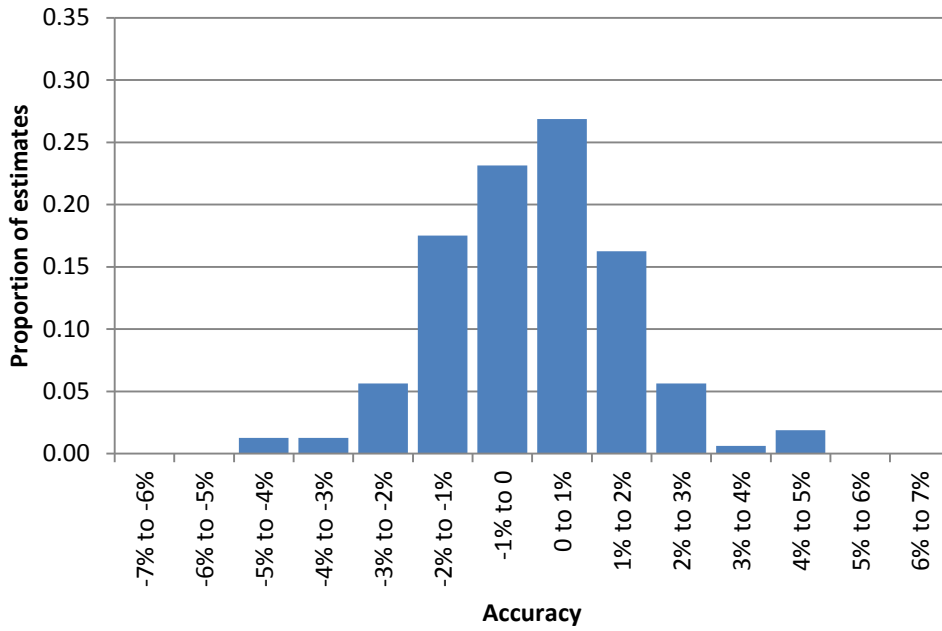

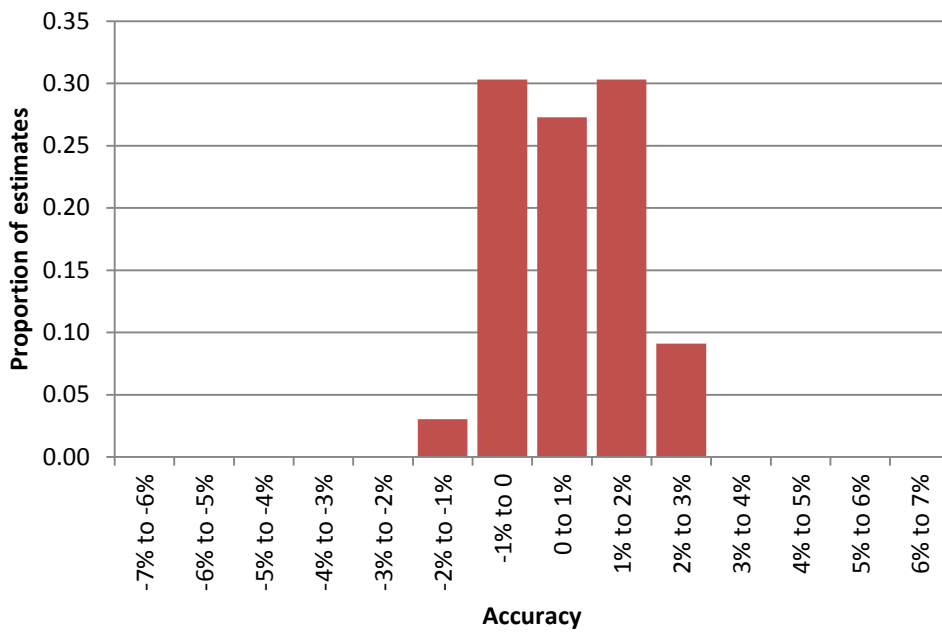## Figure 2: Actual distribution of accuracy of birth predictions using GP registration data

Table 2: Comparison of accuracy of predicted births for year to mid-2015

Local authority	ONS 2015 mid-year birth estimate	Absolute error			Percentage error		
		GP register-based	ONS 2014-based SNPP	GLA 2015 round	GP register-based	ONS 2014-based SNPP	GLA 2015 round
Barking & Dagenham	3,697	-12	-38	15	-0.3%	-1.0%	0.4%
Barnet	5,222	29	158	59	0.6%	3.0%	1.1%
Bexley	3,093	29	-3	-36	0.9%	-0.1%	-1.2%
Brent	5,157	-65	-181	-30	-1.3%	-3.5%	-0.6%
Bromley	4,098	-62	-24	-64	-1.5%	-0.6%	-1.6%
Camden	2,735	19	220	-54	0.7%	8.1%	-2.0%
Croydon	5,705	-99	-197	-63	-1.7%	-3.4%	-1.1%
Ealing	5,389	41	-43	-15	0.8%	-0.8%	-0.3%
Enfield	4,895	78	-25	-6	1.6%	-0.5%	-0.1%
Greenwich	4,547	20	-45	-130	0.4%	-1.0%	-2.9%
Hammersmith & Fulham	2,371	23	161	93	1.0%	6.8%	3.9%
Haringey	4,112	64	38	-9	1.6%	0.9%	-0.2%
Harrow	3,509	-66	-83	68	-1.9%	-2.4%	1.9%
Havering	3,138	-57	-171	-1	-1.8%	-5.5%	0.0%
Hillingdon	4,482	-15	-67	-72	-0.3%	-1.5%	-1.6%
Hounslow	4,339	10	-1	-11	0.2%	0.0%	-0.3%
Islington	2,908	-49	152	-33	-1.7%	5.2%	-1.1%
Kensington & Chelsea	1,823	0	123	-19	0.0%	6.7%	-1.0%
Kingston upon Thames	2,314	9	-4	-122	0.4%	-0.2%	-5.3%
Lambeth	4,549	13	284	45	0.3%	6.3%	1.0%
Lewisham	4,763	85	213	54	1.8%	4.5%	1.1%
Merton	3,330	-23	44	-37	-0.7%	1.3%	-1.1%
Newham	5,998	-138	201	292	-2.3%	3.3%	4.9%
Redbridge	4,801	-31	-190	-186	-0.6%	-4.0%	-3.9%
Richmond upon Thames	2,621	-11	191	71	-0.4%	7.3%	2.7%
Southwark	4,641	-21	371	40	-0.5%	8.0%	0.9%
Sutton	2,822	-34	-152	-173	-1.2%	-5.4%	-6.1%
Tower Hamlets	4,587	-58	281	1	-1.3%	6.1%	0.0%
Waltham Forest	4,670	-60	-38	-133	-1.3%	-0.8%	-2.9%
Wandsworth	5,057	-49	466	174	-1.0%	9.2%	3.4%
Westminster	2,685	-22	208	-93	-0.8%	7.7%	-3.5%
City and Hackney	4,462	-42	191	24	-0.9%	4.3%	0.5%
Greater London	128,520	-466	2,038	-352	-0.4%	1.6%	-0.3%

Source: ONS 2015 mid-year estimates, ONS 2014-based SNPP, GLA 2015 round trend based projections

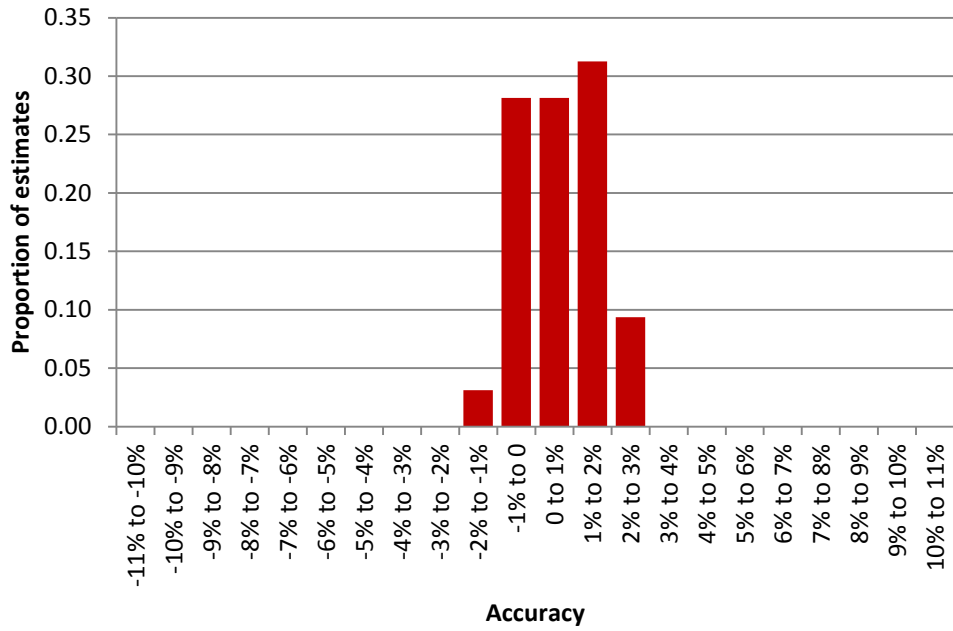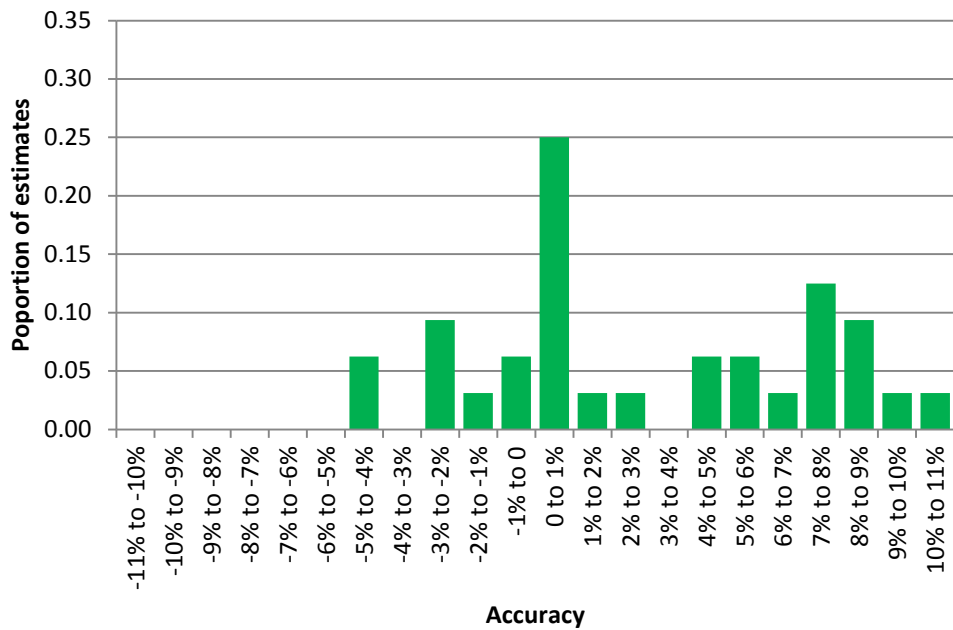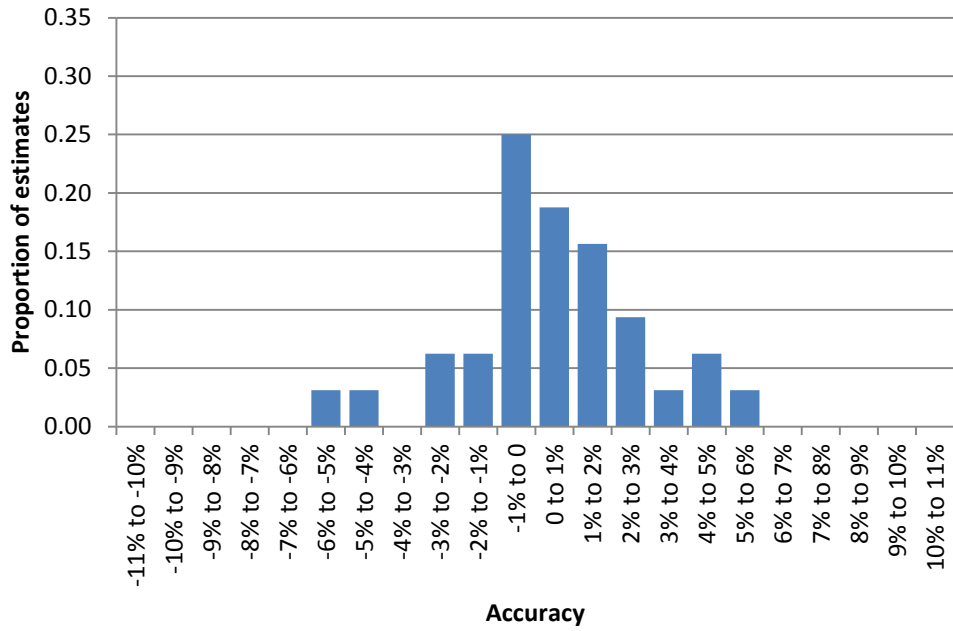**Figure 3: Distribution of accuracy using GP registration data to predict birth estimates**

Figure 4: Distribution of accuracy of ONS subnational projection results
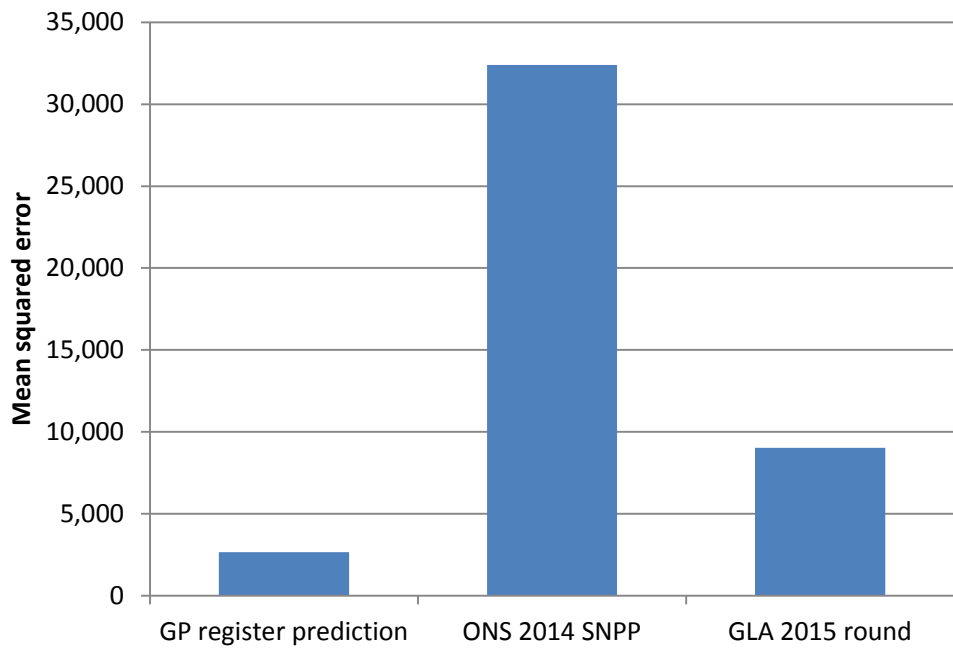

Source: ONS 2014-based subnational population projections

Figure 5: Distribution of accuracy of standard GLA projection results

Source: GLA 2015 round short term trend-based population projections

Figure 6: Mean squared error of mid-2015 births predictions for London local authorities

Predicted births for year to mid-2016

The GLA produced predicted estimates of births for the year to mid-2016 using GP registration data for July 2016. GP data for January 2016 and calendar year births for 2015 were also added to the series used to generate the relationships between births and nought year olds on the GP register. This meant that the ratios defining the relationships between the two quantities were based on the average of eight data points for all local authorities except Hackney and the City of London, for which only six data points were available.

The methodology predicts 129,900 births in London for the year to mid-2016, an increase of 1,400 over the previous year. Figure 7 shows ONS's recent mid- and calendar-year birth estimates for London, together with the set of GP register-based predictions. This illustrates the past accuracy of the approach, and also shows that a rise in births to mid-2016 might also be anticipated based on the calendar year estimate of 129,600.

Table 3 shows the predicted 2016 mid-year birth estimates for each London local authority alongside recent official birth estimates. The results suggest that the largest increases in births between the two periods have taken place in east London, with increases of over 200 in Barking & Dagenham, Newham, Havering, and Hackney. Falls of over 100 births are predicted to have occurred in Ealing, Redbridge, Sutton and Hillingdon.

Figure 7: Annual births in London - estimates and predictions based on GP data
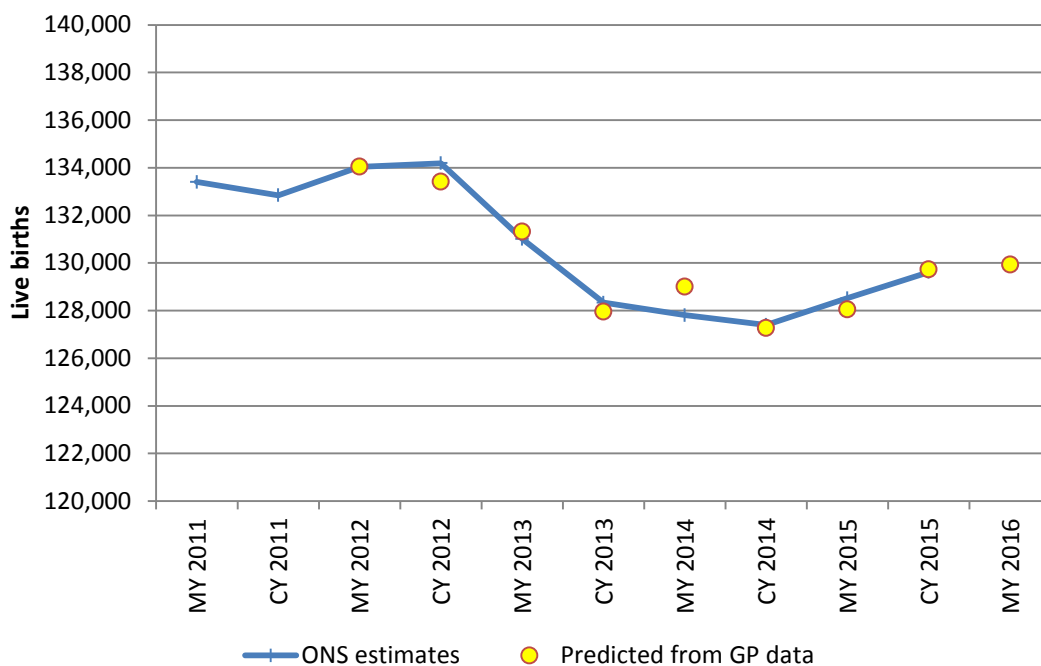
Source: ONS birth summary tables; ONS mid-year estimates; GLA analysis

Table 3: Recent birth estimates and GP register based prediction of mid-2016 birth estimates

Local authority	ONS 2015 mid-year birth estimate	ONS 2015 calendar year birth estimate	Predicted 2016 mid-year birth estimate	Mean ratio age 0 on GP register to births
City of London	63	69	61	0.83
Barking & Dagenham	3,697	3,850	4,002	0.872
Barnet	5,222	5,261	5,267	0.888
Bexley	3,093	3,162	3,118	0.893
Brent	5,157	5,204	5,196	0.864
Bromley	4,098	4,098	4,270	0.911
Camden	2,735	2,699	2,731	0.864
Croydon	5,705	5,833	5,851	0.885
Ealing	5,389	5,210	5,193	0.876
Enfield	4,895	5,027	5,051	0.880
Greenwich	4,547	4,644	4,683	0.862
Hackney	4,399	4,500	4,602	0.865
Hammersmith & Fulham	2,371	2,345	2,403	0.851
Haringey	4,112	4,108	4,134	0.858
Harrow	3,509	3,601	3,588	0.881
Havering	3,138	3,275	3,404	0.889
Hillingdon	4,482	4,394	4,381	0.878
Hounslow	4,339	4,455	4,293	0.890
Islington	2,908	2,939	2,914	0.869
Kensington & Chelsea	1,823	1,805	1,744	0.788
Kingston upon Thames	2,314	2,350	2,336	0.934
Lambeth	4,549	4,620	4,455	0.872
Lewisham	4,763	4,814	4,934	0.887
Merton	3,330	3,412	3,394	0.903
Newham	5,998	6,226	6,303	0.876
Redbridge	4,801	4,798	4,647	0.899
Richmond upon Thames	2,621	2,609	2,552	0.928
Southwark	4,641	4,587	4,563	0.868
Sutton	2,822	2,764	2,708	0.928
Tower Hamlets	4,587	4,560	4,669	0.877
Waltham Forest	4,670	4,651	4,773	0.875
Wandsworth	5,057	5,038	5,023	0.883
Westminster	2,685	2,707	2,689	0.820
Greater London	**128,520**	**129,615**	**129,934**	

Conclusions

The recent release of the 2015 mid-year estimates confirms that GP registration data can be used to predict birth estimates for local authorities with a level of accuracy significantly better than that of established projection methodologies employed by both the ONS and GLA.

The method provides a strong indication of the number of births almost a full year ahead of the release of official estimates. The estimates produced have proven to be sufficiently accurate to recommend their integration into future demographic projections produced by the GLA.

The value of this advanced information will be marginal for many uses, especially those reliant on long-term projections and where overall population is more important than numbers of young children. However, there are applications where the additional timeliness is useful, especially those that rely on near term projections of numbers of young children, as is the case in school place planning.

Ensuring that there are an appropriate number of reception places available for pupils is a particular challenge for planners. Estimates of annual births to mid-year are strong indicators of the likely demand for reception places four years later. Currently there is a one-year lag in birth estimates becoming available, reducing the time available for school place planners to respond to three years. In this case, making accurate estimates available a year earlier gives planners significantly more time to react to the information.

Going forward, the GLA will look to refine the methodology. As a longer series of data becomes available, it may be possible to improve the accuracy of predictions through more detailed analysis of the trends in relationships and by excluding outlier points in the series.

MAYOR OF LONDON