

## Estimating births using GP registration data

January 2016

### Introduction

This note presents analysis that the GLA has carried out looking at the use of General Practitioner (GP) registration data to estimate the number of live births occurring in London.

The desire to find alternative methods for estimating births relates to the timeliness of existing estimates produced by ONS. ONS's birth estimate data is considered to be extremely robust, but estimates of local authority births to mid-year are not published until almost a year after the end of that period. Calendar year birth estimates are timelier, but this data is still released over six months after the end of the collection period. Birth data for lower level geographies typically do not become available until several months after the local authority level estimates.

Birth data is extremely important for local authorities planning provision of primary school places and having an earlier indication of birth numbers would be highly beneficial for local school place planners.

The GLA obtains aggregated extracts of GP registration data for London local authorities on a bi-annual basis – January and July. This data gives numbers of persons registered with a GP by age and ward of residence. The data is a snapshot of the persons on the register at a particular point in time and the GLA receives this data within days of it being collected.

The GLA has investigated whether analysis of the number of children age nought on the GP register at a point in time can be used to provide a useful estimate of the number of children born in an area over the preceding year. If this is the case it would provide an indication of birth numbers far in advance of the release of the official estimates by ONS.

## Methodology

The approach to estimating births examined here involves establishing past relationships between numbers of nought year olds<sup>1</sup> on the GP register and births over the preceding year, and then using these relationships to infer estimates of births based on new GP registration data.

The quality of the estimates generated in this way will be reliant on the consistency of the relationship between birth occurrences and GP registration data from year to year.

The number of births in an area over a year will differ from the number of nought year olds on the GP register at the end of the period for a number of reasons, including:

- delays in children being added to the GP register;
- migration in and out of an area;
- children not being registered with a GP (or registering with private healthcare only);
- infant deaths

With the exception of migration, the effect of each of these factors is to give a lower number of nought year olds on the GP register than births occurring in the preceding year. Migration can drive the relationship either way depending whether there is a net flow of infants to or from the area. The consistency of the relationship between the two sets of data will depend upon the consistency of each of the factors listed above.

This note examines the past relationships between counts from the GP register and annual birth estimates and then looks at the results of using these relationships to estimate births in London.

The analysis carried out was undertaken using both borough- and ward-level birth and GP registration data, allowing comparison to be made between the results at different geographic levels.

### Past relationships between GP registrations and birth estimates

The GLA holds snapshots of counts from the GP register at six-month intervals from July 2012-onwards. The dates of these snapshots are as follows:

- 1 July 2012;
- 1 January 2013;
- 1 July 2013;
- 1 January 2014;
- 8 July 2014;
- 8 January 2015; and
- 8 July 2015

For this exercise, January counts were compared with calendar year birth data from ONS Vital Statistics releases. July counts were compared with births to mid-year. Mid-year births for local authorities were taken from the annual mid-year estimate series. Ward births to mid-year were provided by ONS on request.

For each local authority or ward, the ratio of nought year olds to annual births was calculated. This exercise was repeated for each set of GP data. As ward births to mid-2014 are not yet available, there are five sets of local authority-level ratios but only four sets of ward-level results.

---

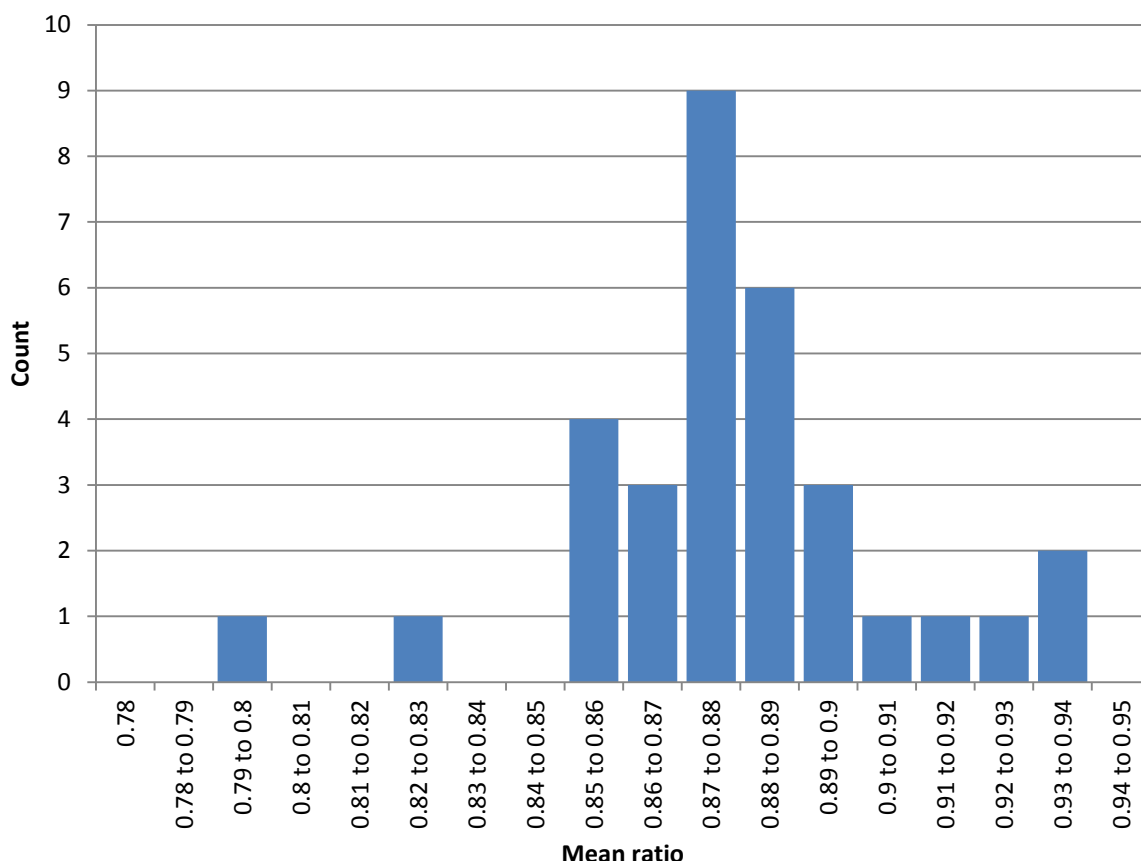
<sup>1</sup> Nought year olds being persons yet to have reached their first birthday.

### Local authority level

In this analysis, data for City of London and Hackney have been combined and a single ratio calculated. This has been done for consistency as the calendar year birth data is received in this form. The analysis therefore considers 32 areas in London.

Figure 1 and Table 1 show the distribution of the mean ratios calculated for each local authority. Between local authorities, the ratio varies between 0.791 and 0.932 with a mean value of 0.879.

**Figure 1: Mean ratio of age 0 on GP register to births in preceding year**



Source: NHS Connecting for Health, ONS birth estimates

Figure 2 shows the geographical distribution of mean ratios across London calculated from the five sets of local authority-level ratios. Kensington & Chelsea and Westminster in Central London have the lowest mean ratios. These are boroughs which are likely to have a higher number of nought year olds registered with private GPs. Such areas of central London are also likely to see a net outflow of families. Both these factors would contribute to the number of nought year olds being lower than the number of births.

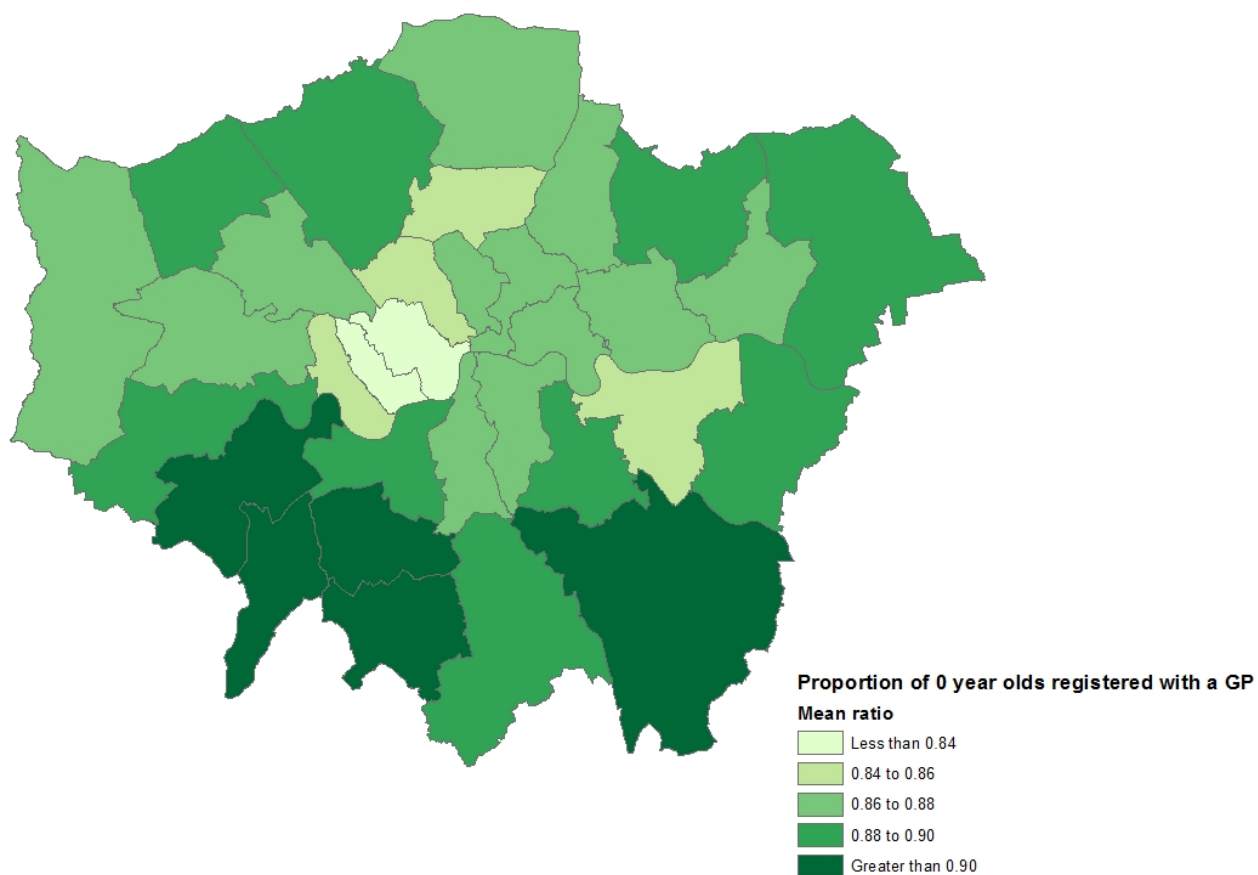
Outer London boroughs have higher mean ratios than those in inner London, with those in the south (Kingston upon Thames, Richmond upon Thames, Sutton, Merton and Hounslow) having the very highest ratios. These boroughs are likely to experience a net inflow of families.

**Table 1: Mean ratio of age 0 on GP register to births in preceding year by local authority**

Local authority	Sample mean
Barking & Dagenham	0.873
Barnet	0.884
Bexley	0.891
Brent	0.866
Bromley	0.916
Camden	0.860
Croydon	0.889
Ealing	0.874
Enfield	0.875
Greenwich	0.858
Hammersmith & Fulham	0.850
Haringey	0.851
Harrow	0.886
Havering	0.893
Hillingdon	0.879
Hounslow	0.889
Islington	0.870
Kensington & Chelsea	0.791
Kingston upon Thames	0.932
Lambeth	0.873
Lewisham	0.882
Merton	0.905
Newham	0.880
Redbridge	0.900
Richmond upon Thames	0.932
Southwark	0.868
Sutton	0.928
Tower Hamlets	0.879
Waltham Forest	0.879
Wandsworth	0.888
Westminster	0.826
Hackney & City of London	0.861

Source: NHS Connecting for Health, ONS birth estimates

**Figure 2: Mean ratio of age 0 on GP register to births in preceding year**



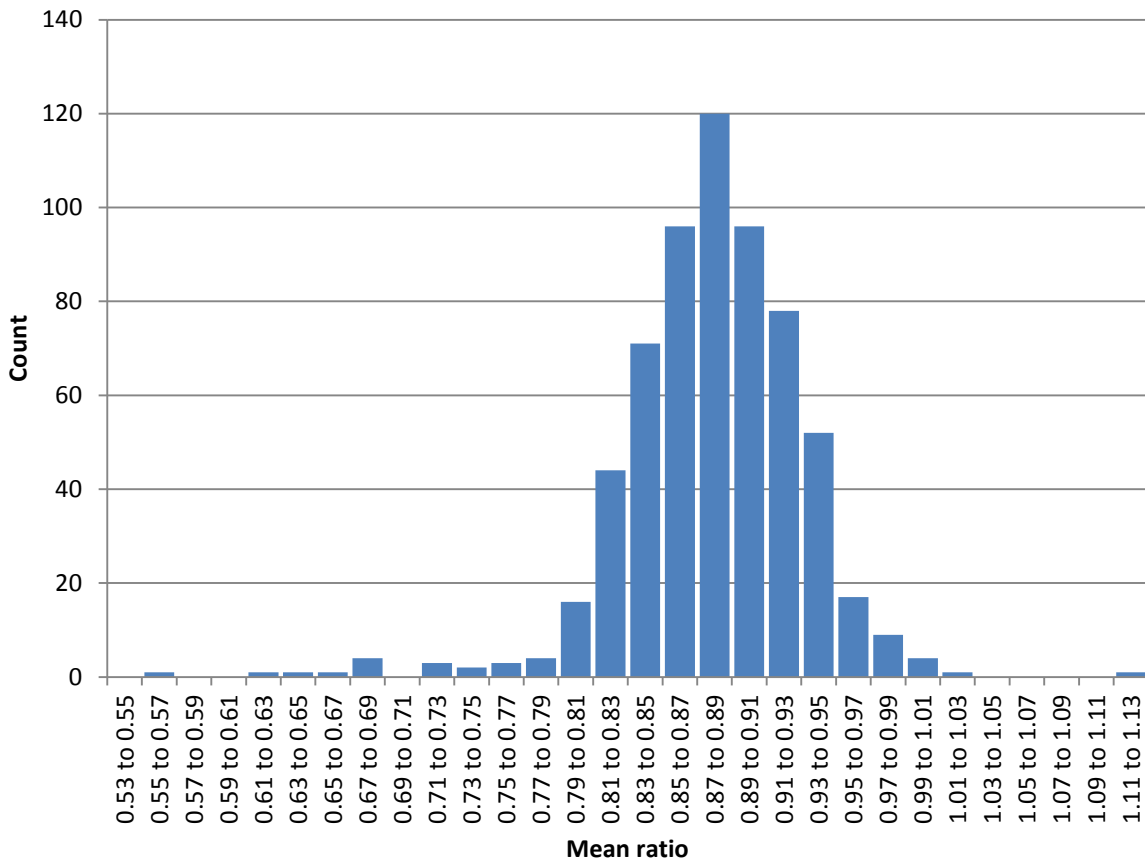
Source: NHS Connecting for Health, ONS birth estimates

### Ward level

Data for the City of London as a whole was included as a single area alongside wards (as defined in 2011 and prior to subsequent boundary changes in some local authorities). This gave a total of 625 areas for the analysis.

Ratios of nought year olds on the GP register to births in the preceding year were created as in the local authority analysis. The distribution of the mean of these results for each area is shown as a histogram in Figure 3.

**Figure 3: Mean ratio of persons age 0 on GP register to births in preceding year**

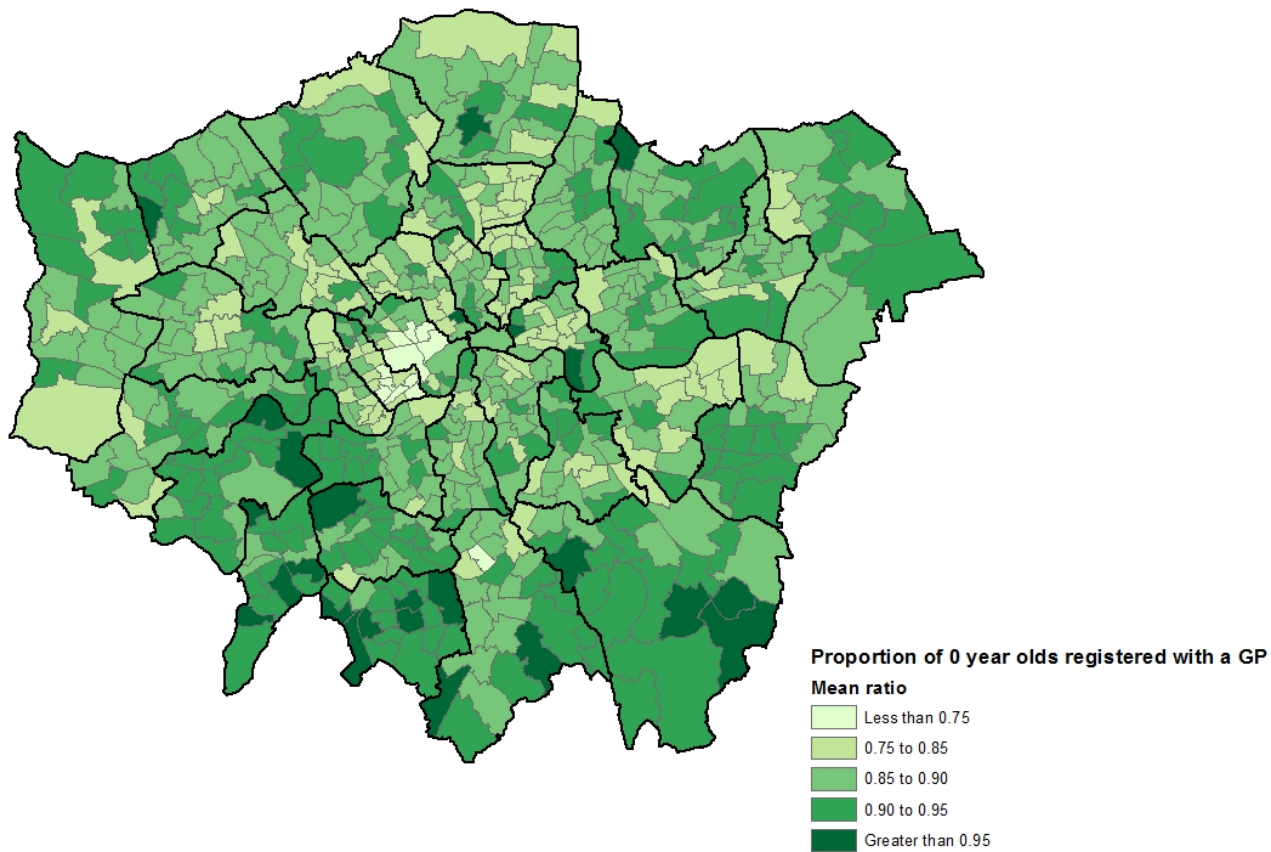


Source: NHS Connecting for Health, ONS birth estimates

The range of ratios for wards (Figure 3) is much greater than for boroughs (Figure 1), with the ratios varying between 0.62 and 1.07 (with a mean value of 0.88). Two wards (St James in Kingston upon Thames, and Selsdon and Ballards in Croydon) have mean ratios greater than one. This indicates that there are a greater number of nought year olds registered with a GP than the number of births in the area over the preceding year. These are areas which are likely to have net inflows of nought year olds resulting in higher numbers of this age when compared with the number of births.

Figure 4 shows the geographic distribution of the mean ratios across London at ward-level. There are cluster of wards with low mean ratios in parts of Kensington & Chelsea and Westminster. Wards with high mean ratios are largely spread throughout boroughs in the south, with clusters in Orpington in Bromley, Wimbledon Village in Merton, and Cheam, Carshalton and Beddington in Sutton.

Figure 4: Mean ratio of age 0 on GP register to births in preceding year



Source: NHS Connecting for Health, ONS birth estimates

## Investigating likely accuracy of predictions based on GP register data

To generate estimates of births based on the GP register data, the number of nought year olds on the register is divided by a ratio of nought year olds to births based on historic data.

$$\text{Estimated births} = \frac{\text{Number age 0 on GP register}}{\text{Historic ratio of Births to Number age 0 on GP register}}$$

To gain an understanding the level of variability in results that might be expected using this approach in the future, birth estimates corresponding to each set of GP register data were produced. In each case, the historic ratio used was the mean of the ratios calculated for each set of GP data *excluding* the set being used to create the estimate. The difference in these estimates to ONS birth estimates gives an indication of likely future accuracy of the results.

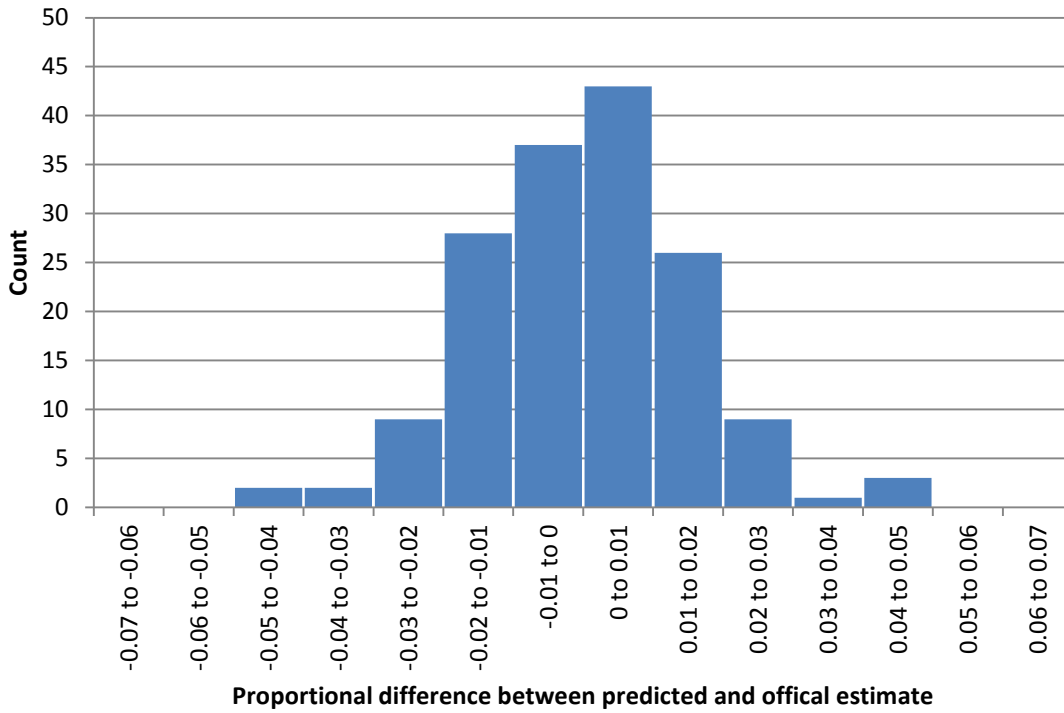
Five estimates were generated for each local authority giving a total of 160 estimates. These were compared with ONS birth estimates and the proportional difference calculated. The distribution of these proportional differences is plotted in Figure 5.

50 per cent of the predictions lay within one per cent of the actual estimates, 84 per cent within two per cent of the actuals, and 95 percent within three per cent of the actuals.

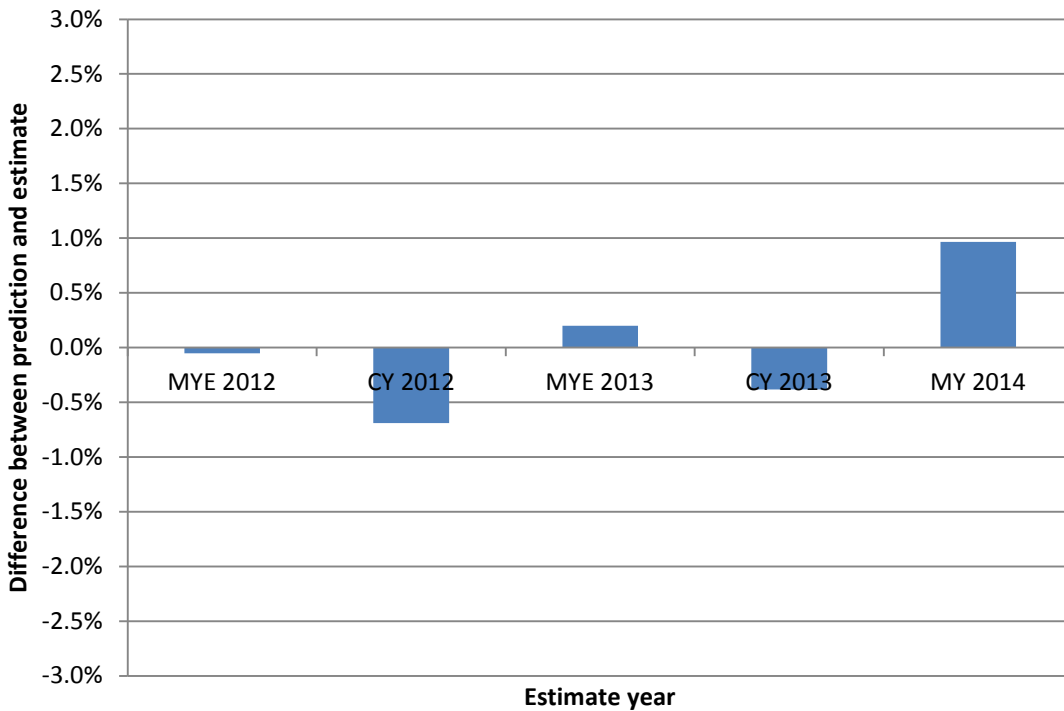
The accuracy of predictions for London as a whole is shown in Figure 6. At this level, all predictions were within one percent of the actual estimates, with a range of 0.05 per cent (difference of 71) to 1.0 per cent (difference of 1,232). The mean error was 0.5 per cent.



**Figure 5: Accuracy of predicted local authority births based on GP register data**



**Figure 6: Accuracy of predicted London births based on GP register data**



## Comparison with predictions based on calendar year births

To give some context to the results of the GP register-based approach, comparison can be made with alternative approaches. In recent years, the GLA has used calendar year data (released in August each year) as a predictor of births to the next mid-year (released ten months later, in June) – typically using them as a direct proxy.

The rationale for using calendar year births as a predictor is that the second half of the calendar year period is common with the first half of the mid-year period. Prior internal analysis showed this approach to give better projection accuracy than approaches relying only on the most recently available births to mid-year.

Proportional difference between the calendar year ‘predictions’ and birth estimate ‘actuals’ were calculated for each set of mid-year estimates birth estimates from 2001-02 to 2013-14. These differences were calculated for local authorities and London as a whole (Figures 7 and 8).

416 results were produced for local authorities (13 years x 32 authorities). 29 per cent of predictions lay within one per cent of the actuals, 60 per cent lay within two per cent, 81 per cent lay within three percent, and 96 per cent lay within five percent.

For London, the difference between predicted and actual values ranged from 0.02 per cent (difference of 23) to 2.6 per cent (difference of 2,804), with a mean error of 1.2 per cent.

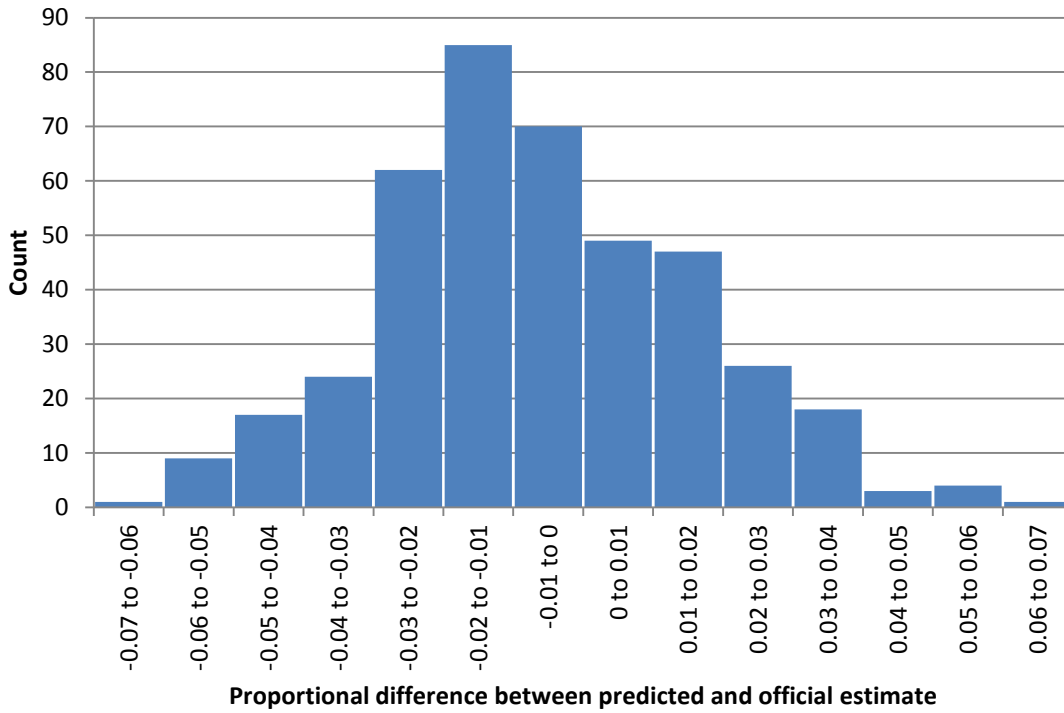
There is a clear pattern to the accuracy of the approach for London. The method under predicted for all mid-year estimates from 2002 to 2012 except for 2009, and over predicted for 2013 and 2014.

This pattern is easily understood if considered alongside the annual birth trends for the period (Figure 9). Under-prediction occurred when births were rising year-on-year and over-prediction when births were falling. The 2008 calendar year estimate was an accurate predictor of births to mid-2009 - this coincides with a pause in the trend of rising births.

The tendency to under predict when births are rising and over predict when births are falling is a feature of the approach of using calendar year data as a proxy. Accuracy of predictions is a function of the size of the year-on-year change in annual births: larger changes in birth numbers lead to lower prediction accuracy.

By contrast, the accuracy of predictions made using GP register data should be largely independent of the size of the rate of change in the number of births. Accuracy is instead dependent on the stability of relationships between children on the GP register and the number of births in the preceding year. These relationships are sensitive to changes in: patterns of migration; date that register data is collected; proportion of children on the GP register.

**Figure 7: Accuracy of predicted local authority births based on calendar year data**



**Figure 8: Accuracy of predicted London births based on calendar year data**

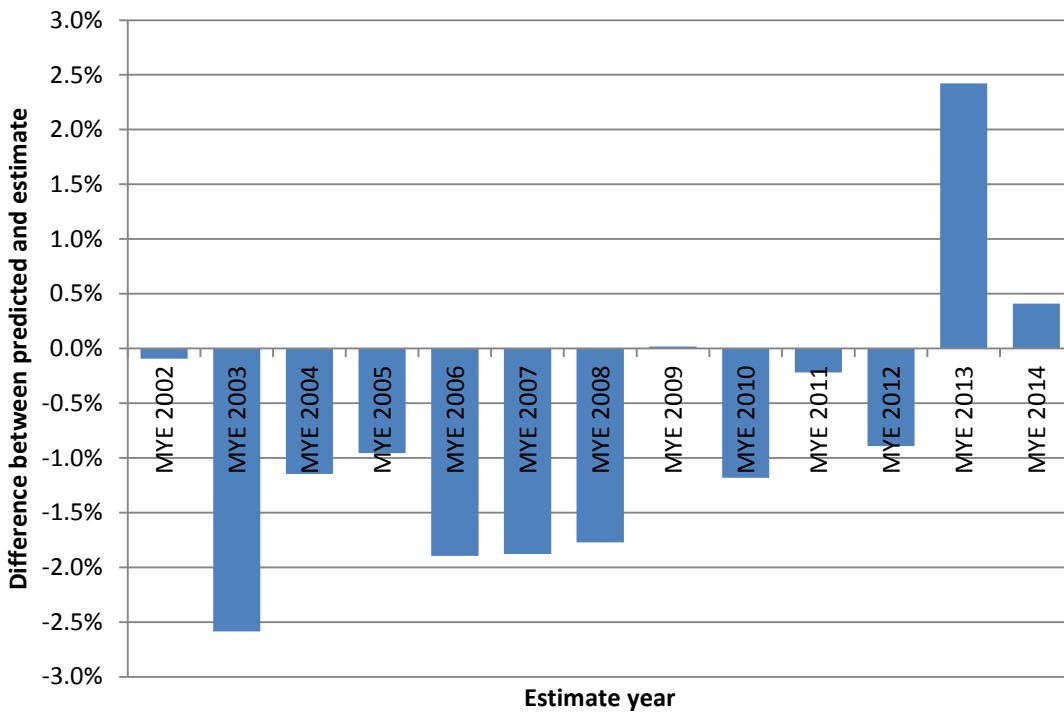
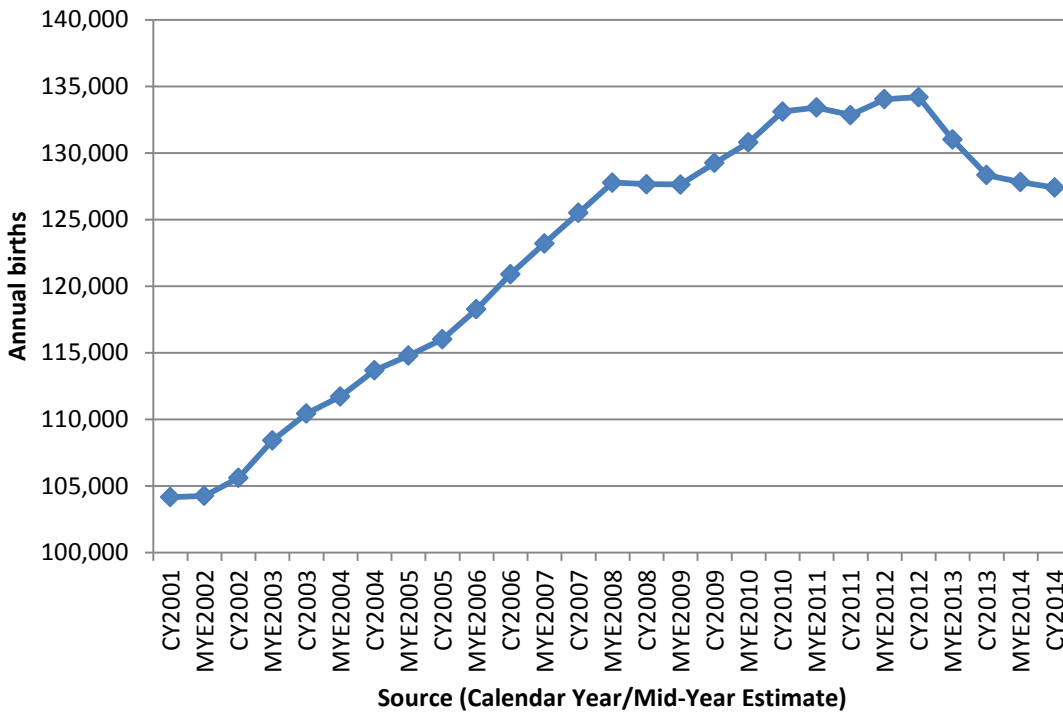


Figure 9: Annual births for London 2001 to 2014



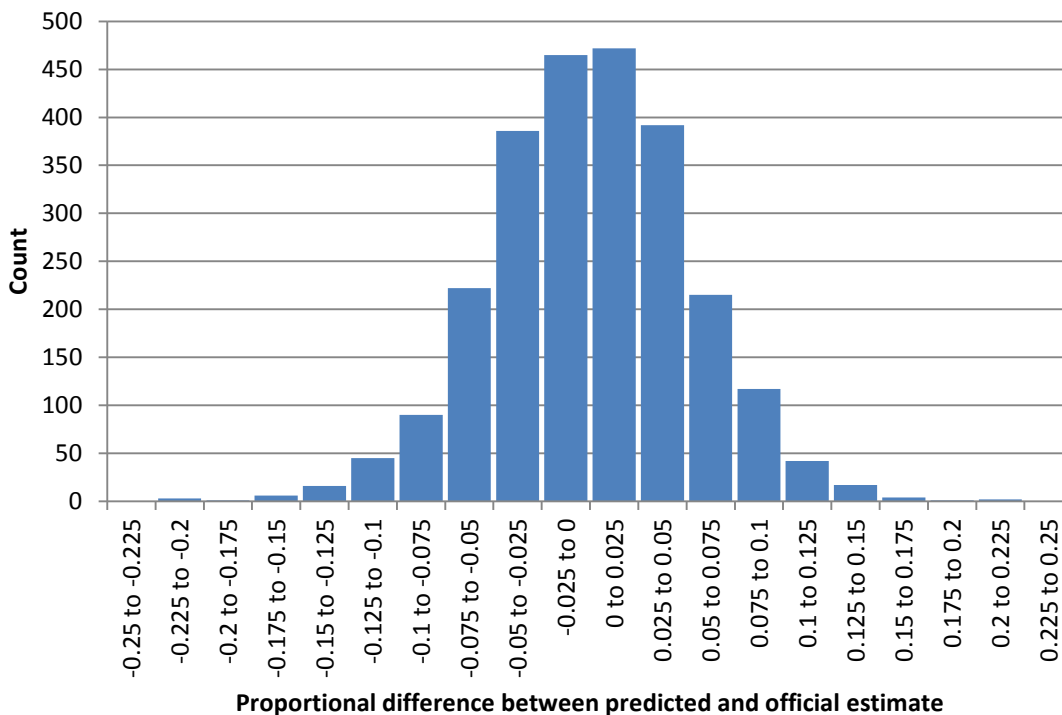
ONS mid year estimates, calendar year birth estimates

### Ward level accuracy

Four estimates were produced for each ward (plus City of London) giving a total of 2,500 estimates. The distribution of the proportional differences of these estimates relative to the actuals is shown in Figure 10. The spread of results is much larger than was the case at local authority level. This is due to the added variability resulting from the smaller numbers involved.

At this smaller geography, accuracy was lower than at local authority level. 69 per cent of predictions fell within five per cent of the actuals, and 94 per cent fell within ten per cent.

**Figure 10: Differences between predicted ward birth estimates and actuals**



### Predicting 2015 mid-year estimate births

The GLA has received GP registration data for July 2015. By applying the methodology described above this data was used to create predictions of local authority birth estimates for the year to mid-2015. These predicted birth estimates are shown in Table 2, presented alongside the 2014 birth estimates for comparison. Official estimates of births to mid-2015 are scheduled for release in June 2016.

Based on an overall count of 113 thousand children age zero on the GP register, the approach predicts an estimate of 128 thousand births for the year to mid-2015. This would be a modest rise of 250 births over the estimate to mid-2014. This aggregate London figure hides a wider range of predicted change in births. The data implies a fall of 425 in annual births in Newham and a rise of 178 births in Redbridge.

The decision was made not to produce ward-level birth predications due to the small number of sets of past ratio data currently available. Further work is planned to investigate the value of producing predictions at lower geographies once 2014 calendar and mid-year birth estimates are available by ward.

**Table 2: Predicted births for year to mid-2015 based on GP registration data**

	Mean ratio age 0 on GP register to births	Count of age 0 on July 2015 GP register	Predicted 2015 mid-year birth estimate	ONS mid-2014 birth estimate
Barking & Dagenham	0.8735	3,219	3,685	3,699
Barnet	0.8842	4,643	5,251	5,275
Bexley	0.8911	2,782	3,122	3,019
Brent	0.8663	4,411	5,092	5,160
Bromley	0.9158	3,696	4,036	4,002
Camden	0.8599	2,368	2,754	2,679
Croydon	0.8885	4,981	5,606	5,648
Ealing	0.8739	4,745	5,430	5,462
Enfield	0.8752	4,352	4,973	4,876
Greenwich	0.8576	3,917	4,567	4,405
Hammersmith & Fulham	0.8504	2,036	2,394	2,494
Haringey	0.8508	3,553	4,176	4,091
Harrow	0.8862	3,051	3,443	3,574
Havering	0.8928	2,751	3,081	3,088
Hillingdon	0.8787	3,925	4,467	4,369
Hounslow	0.8892	3,867	4,349	4,344
Islington	0.8700	2,487	2,859	2,841
Kensington & Chelsea	0.7909	1,442	1,823	1,820
Kingston upon Thames	0.9317	2,164	2,323	2,169
Lambeth	0.8730	3,983	4,562	4,571
Lewisham	0.8821	4,276	4,848	4,753
Merton	0.9046	2,992	3,307	3,308
Newham	0.8795	5,154	5,860	6,285
Redbridge	0.8997	4,292	4,770	4,592
Richmond upon Thames	0.9319	2,432	2,610	2,698
Southwark	0.8684	4,012	4,620	4,663
Sutton	0.9277	2,586	2,788	2,642
Tower Hamlets	0.8785	3,979	4,529	4,520
Waltham Forest	0.8793	4,054	4,610	4,540
Wandsworth	0.8878	4,446	5,008	5,206
Westminster	0.8258	2,199	2,663	2,574
Hackney & City of London	0.8606	3,828	4,448	4,440
Greater London		112,623	128,054	127,807

## Conclusions

Analysis of past relationships between GP registration and birth estimates data indicates that GP registrations can be used to predict birth estimates by local authority with a good level of accuracy and up to eleven months ahead of the release of official data.

The ability to gain an early indication of birth estimates has the potential to improve the timeliness of the data included in population and school roll projections.

Recent rounds of the GLA's local authority level projections have used calendar year birth data as a proxy for subsequent mid-year estimates. Analysis of the accuracy of using GP register or calendar year birth data to predict mid-year births suggests that the approach of using GP register data is generally more robust, with 95 per cent of local authority predictions lying within three per cent of actual estimates. Accuracy of the approach at small geographical levels is less good, with 94 per cent of ward-level predictions lying within ten per cent of actuals.

Ahead of issuing the next round of population projections, the GLA will review the options for projecting fertility and births to determine whether to make use of predictions based on the methodology described in this note.

The approach has been applied to GP registration data for July 2015 to produce predictions of births for the year to mid-2015. These results suggest that the total number of births in London over this period were similar to the number for the year to mid-2014, though significant changes occurred within individual local authorities.