Investigation into the ONS Mid-Year Population Estimates

1. Introduction

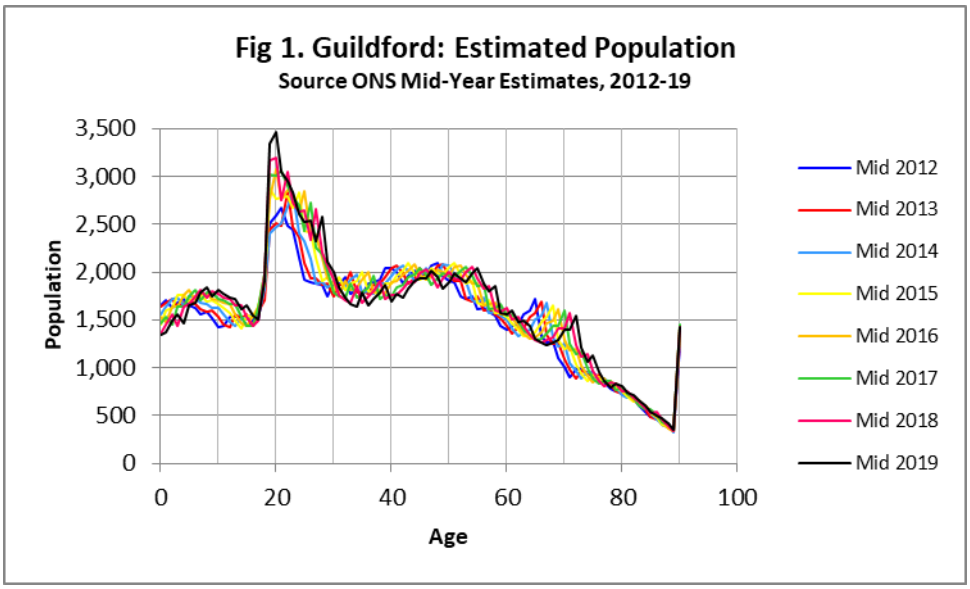
The purpose of this document is to report some apparent anomalies in the Office for National Statistics (ONS) population data that underlies the Guildford Local Plan. Two separate ONS datasets were used to underpin the Local Plan:

- The mid-year population estimates (MYEs), which provide an annual "snapshot" of the estimated population for each local authority in the UK broken down by each year of age for both males and females.
- The sub-national population projections (SNPP), that are published every two years and which provide the projected population figures approximately 20 years into the future, broken down by the components of change, namely births, deaths, and inward and outward migration (both internally within the UK, and internationally).

This paper is not intended to replay the considerable degree of attention focussed on Guildford's future population trajectory during the development of the Local Plan (which mainly examined the SNPP data and the way in which it was interpreted and developed in the Strategic Housing Market Analysis (SHMA)). Instead, it focuses simply on the reliability of the fundamental MYE population data that underlies both the SNPP and the SHMA.

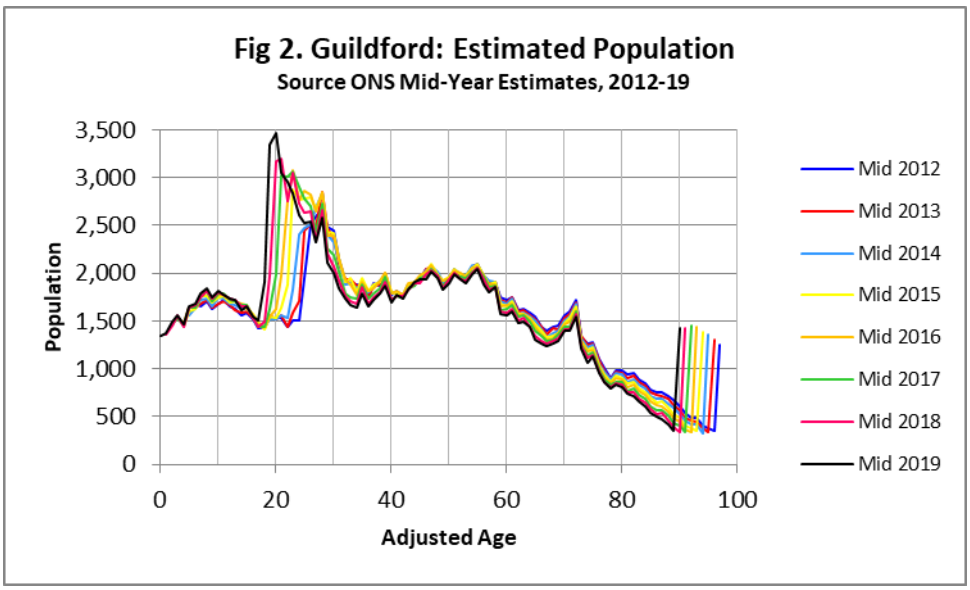
2. Review of the ONS Mid-Year Population Estimates for Guildford

Figure 1 presents MYE data for 2012 to 2019, and shows the ONS view of the age structure of Guildford's population.

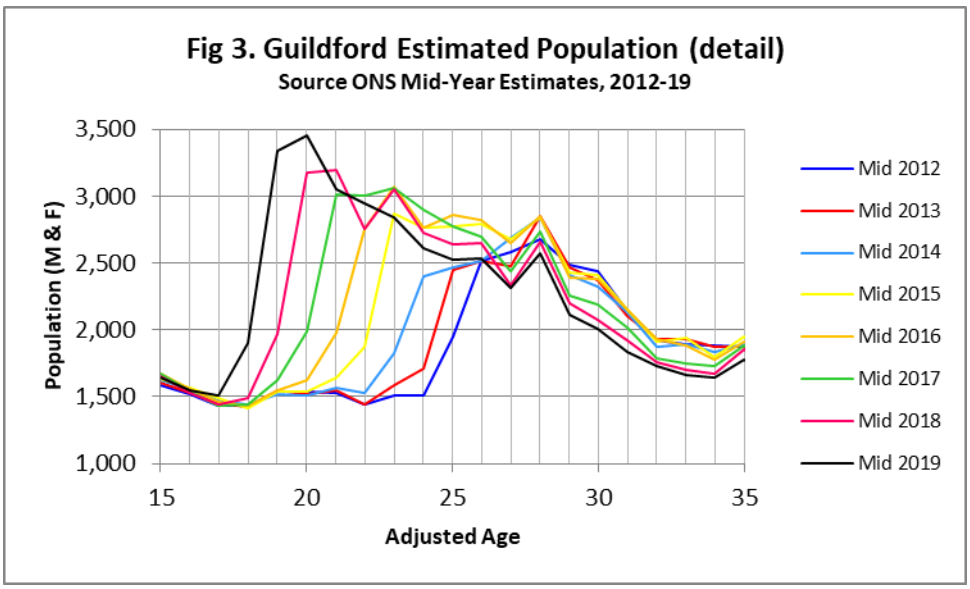


In each subsequent set of MYE data, the population ages by one year, and any identifiable features population structure consequently shift to the right by a year. The sharp rise in the curves on the right of the figure is because ages of 90 and more are grouped into a single data point in the ONS data. The other main feature is the prominent peak at about age 19; this represents a significant influx of students moving to Guildford for higher education. However, in Figure 1 the overlapping curves tends to obscure one another, and it a clearer view of the annual changes can be obtained by plotting population data against adjusted ages, as in Figure 2. (The adjusted age appropriate to any particular

MYE dataset is obtained by reducing true ages by the interval between 2012 and the year to which the dataset applies. For example the adjustment for MYE 2015 data would be an age reduction of 3 years, ie. 2015 - 2012.)

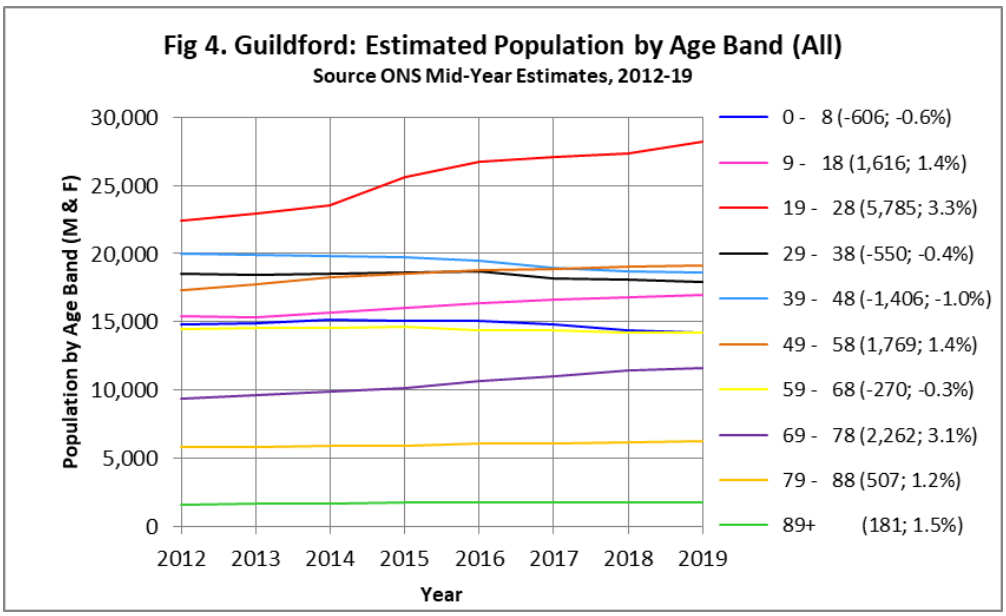


The main features of the "student bulge" are that with each succeeding year, both the height and the width of the bulge increases, and that the later annual bulges almost fully enclose the earlier ones. This can be seen more clearly in Figure 3 which presents an enlarged version of the same data as that in Figure 2. The ONS data clearly implies that between 2012 and 2019 there were sharply increasing numbers students migrating to Guildford, and that year-by-year those students stayed in Guildford for longer.



The extent of this apparent growth over the 10-year age band that includes most of the "student bulge" (ie. true ages of 19 and 28) is very considerable. Figure 4 presents the data for a set of 10-year age bands for the complete population (except for the extreme bands for which the age band widths are curtailed to match the available data, ie. 9 years from 0 to 8 inclusive, and a band of indeterminate width for the 89+ age band). The legend to the right of Figure 4 also includes the actual change in population for that band, and the corresponding annual compound rate of growth or decline. The MYE data estimated the growth of the total Guildford population between 2012 and 2019 to be 9,288

(0.9%), whereas the growth in the 19 to 28 age band was 5,785 (3.3%). So the population growth in the "student bulge" age band of 19-28 accounted for nearly two thirds of the total estimated population growth in the MYE data (62.3%).



These figures for the "student bulge" (ie. a 3.3% annual rate of growth, and its status as the principal driver of population growth in Guildford) seem remarkably high. The clear interpretation of the ONS data presented in Figure 3 is that in the 7 years between 2012 and 2019 the following two changes occurred:

- Approximately 2,000 young people a year were moving to Guildford in 2019, compared with only about 1,100 just five years earlier in 2012.
- For each year in the period 2012 to 2019, the duration for which young people stay in Guildford is a year longer than the duration in the previous year's ONS MYE estimate.

The first of these changes seems doubtful at best, but in principle could be checked against HESA data. The second change would appear to be very unlikely, and it would seem to be far more likely to be caused by a systematic error in the methodology used to derive the estimates.

In view of the doubtful reliability of the MYE estimates an independent "sanity check" was undertaken. Given that the width of the "student bulge" extends well into the principal child-bearing ages (especially for the later MYEs), the population trend of females of those ages was compared against the corresponding trend of infants aged less than 1 year old. It is important to note that the data required to make this comparison are all drawn for the same series of annual MYE datasets so they should be mutually consistent. Although a number of potentially confounding factors might influence this comparison, it was nevertheless anticipated that some broad similarity between these two trend lines would be observed. The age band of females for this sanity check was chosen as between 24 and 31 years of age; 24 as a lower age limit excludes students aged 23 years of age or less (corresponding to the age when most undergraduates would have completed their degree), and 31 represents the greatest age of the "student bulge" in the 2019 MYE data (see Figure 1). According to Table 3 of current ONS data[1], the 24-31 age band includes the mothers of approaching half of live

[1] Births by Parents' Characteristics, England and Wales (https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsbyparentscharacteristics

births in 2019, so it provides confidence that 24-31 band for Guildford would include a sufficiently large sample of births from which to draw valid comparisons between the two trend lines.

It should be noted that the check was carried out without any expectation that the **number** of infants could be accurately inferred from the number of females in the relevant age band. The purpose of the check was to examine whether that **trend of infant numbers** was broadly consistent with the corresponding **trend of the number of females** in the selected age band. Accordingly, the data presented in Figure 5 have all been normalised with respect to 2012 values to provide a ready means of comparison. 2012 was chosen as the baseline because it was the first year from which MYEs were readily available after the 2011 Census.

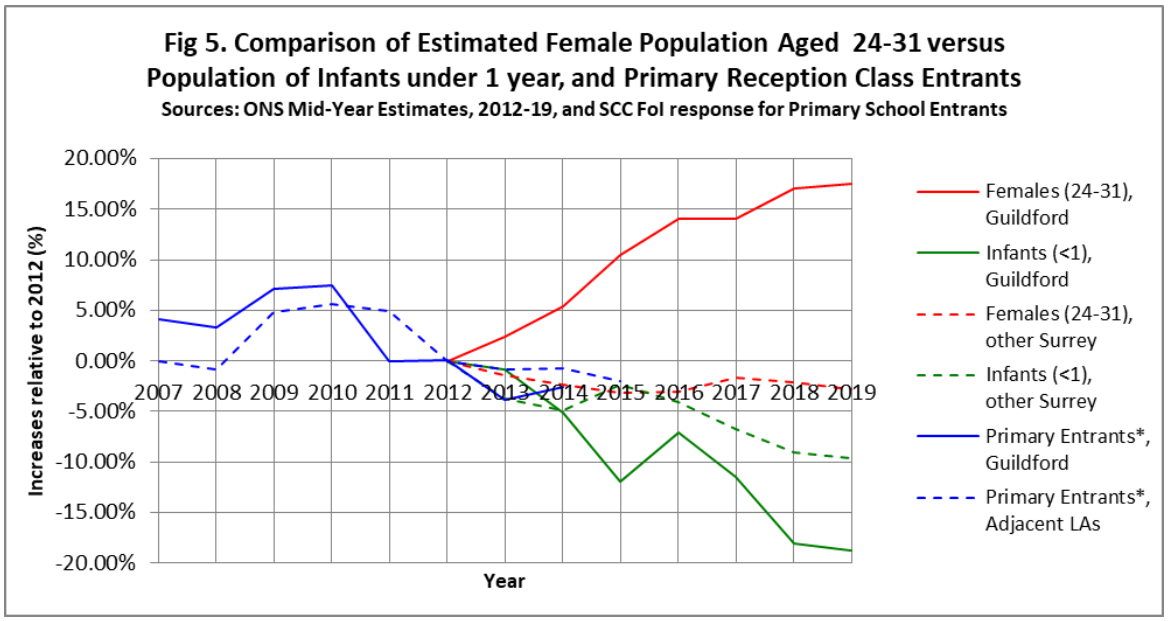


Figure 5 show a very large divergence between the solid red curve for Guildford for females between 24 and 31 years of age versus the solid green curve for infants less than one year of age. In contrast, the same information for all other Surrey boroughs (shown in the same colours but with dotted lines) displays very similar trend lines between 2012 and 2016, and only modest divergence thereafter. The discrepancy between the two Guildford trend lines therefore strongly suggests that at least one of them is likely to be incorrect.

Finally, as an additional element of the sanity check, historic data on the number of Primary School reception class pupils in Guildford and its adjacent local authorities (Elmbridge, Mole Valley, Waverley and Woking) was obtained from the local education authority (Surrey County Council) for the period 2012 to 2020 (which was all that was available). This data (with five years deducted from the dates to provide a proxy for the number of births five years earlier) is also shown on Figure 5 as the blue curves. The Figure shows that the trend of these curves is consistent with MYE data for infants aged less than one year. The only curve in Figure 5 that exhibits a trend significantly different from any of the others is the one relating to females in Guildford aged between 24 and 31 years of age. Taken together with the clear evidence presented graphically in Figure 3, this raises significant doubts on the reliability of the MYE estimates in the 24-31 year age band.

3. Discussion

Section 2 above presents the results of an investigation into the ONS MYE data for the years 2012 to 2019, and identified the following:

- a specific problem of increasing population in the 19-28 year age band (see Figure 3) – both in terms of the number of young people coming to Guildford each year, and in the apparent duration of their residence.
- a large discrepancy between the trend line for females in the 24-31 age band versus the trend line for infants less than only year old (see Figure 5).

Specifically, the "student bulge" becomes progressively larger year by year, and the MYE estimates imply that very few students leave Guildford after a typical 3 or 4 year undergraduate course; in fact in any given MYE dataset, the duration for which young people stay in Guildford is a year longer than the duration in the previous year's MYE dataset. Figure 4 above showed that there was an implausibly large rate of population growth (3.3%) per annum in the 19-28 age band between MYE 2012 and MYE 2019. If – as this paper suggests – there is a procedural error in the MYE estimates of student numbers, the extent of that error could be estimated by assuming that the "student bulge" in fact remained of constant size instead of growing year by year. On this basis, the error over the 7-year period between 2012 and 2019 would then be simply the difference between the population estimates for that age band for those two years, ie. 5,785 (as shown in Figure 4).

Any error in the procedure for producing the MYE population estimates is typically corrected at a National Census, but if it remains uncorrected it will continue to propagate in each succeeding annual MYE until it is once again corrected at the next Census. Therefore a rough estimate of the error over the decade between the 2011 and the 2021 Censuses is simply 8,264 (ie. 5,785 x 10 / 7). It is not the first time that an error of this general magnitude has occurred; Section 3.10 of the "West Surrey Strategic Housing Market Assessment – Guildford Addendum Report 2017 (Final)" stated:

> ***"Unattributable Population Change***
>
> *UPC is an adjustment made by ONS to reflect differences essentially between its 'components of change data' (births, deaths and migration estimates) and what Census data in 2001 and 2011 showed regarding population growth. It thus relates to the 2001-11 period. In Guildford UPC is positive over the 2001-11 period and totals 7,173 persons, a not insignificant amount."*

By definition, the ultimate cause of UPC is not known. However, the ONS publication "Mid-year Population Estimates – Quality and Methodology Information"[2] discusses a number of potential areas of difficulty in making such estimates, such as the problem in tracking UK students who complete higher education courses and move to other locations without registering with a GP, and the difficulty of tracking international students who leave the UK after their course by using the International Passenger Survey. Difficulties in capturing the relevant data in both these areas would produce exactly the effects highlighted in Figures 1 – 3 above. In addition, datasets that produce perfectly acceptable results at national levels frequently begin to exhibit progressively less reliable results as they are they are focused down to smaller sub-national geographic areas. Finally in the more recent MYE datasets, it is far from clear that the effects of Brexit have been fully taken into account.

---

2 https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/annualmidyearpopulationestimatesqmi

Government policy was for Guildford's Local Plan to be grounded on 2014-based ONS population and housing data. Although that baseline MYE 2014 data might be expected overstate the population by only about 30% of the 8,264 figure above (because 2014 lies only 30% of the way through the 10 year period between Censuses), the forward projections by ONS (population – SNPP) and by MHCLG (housing projections) would nevertheless propagate the error trends forward at a similar rate. To a first order of magnitude, any errors such as those discussed in this paper could therefore overstate Guildford's population by as much as about 12,000 over the 15-year Local Plan period. This would equate to roughly 5,000 dwellings, and would certainly invalidate the basis of the Local Plan adopted in 2019.

4. Conclusions and Recommendations

1. There are very good reasons to challenge the accuracy of the ONS MYE population estimates for Guildford.

2. The likely errors in the ONS MYE population estimates for Guildford are sufficiently large to invalidate the present Guildford Local Plan adopted in 2019.

Consequently, the following courses of action are recommended:

3. GBC should engage actively as a major participant in all follow-up work arising from the review of the ONS MYE estimates currently being undertaken by the UK Statistics Authority and the Office of Statistics Regulation (UKSA/OSR)[3]. (That review focuses particularly on moderate-sized towns with substantial universities.)

4. GBC should seek support from UKSA/OSR for an early release of 2021 Census data on the Guildford population (as the quickest and most effective way of confirming the work reported in this paper).

5. GBC should include the detailed review of the fundamental population and housing data in the more general review work arising from the resolution[4] passed at the Council Meeting on 13th April 2021,

6. GBC should take immediate legal advice to determine whether credible concerns about the reliability of the ONS MYE data underpinning the current Local Plan can legally be considered by the Planning Committee when determining planning applications.

D J Reeve
8th May 2021
All recipients of this paper (both direct and indirect) are free to distribute it as they see fit.

[3] The present UKSA/OSR review was undertaken as a result of a considerable body work carried out by residents in Coventry in connection with their Local Plan. They encountered very similar (but even more extreme) issues with the ONS MYE data as those described in this paper. I am grateful for their work, and am pleased to acknowledge it.

[4] "This Council therefore: RESOLVES: To continue to plan the review of the Local Plan and evidence base, to obtain expert independent assessment of the new issues and changed circumstances, to advise on the appropriate route to review the Local Plan, update the Plan's Evidence Base and then update the Local Plan in order to secure the best outcomes for our community and borough."