**Using demographic analysis in producing the best possible population statistics for England and Wales – Options around a national adjustment.**

**Executive Summary, purpose of this paper and who needs to be involved**

Demographic Analysis focuses on expected patterns given the way that population accounting works and expected trends, for example we would not expect mortality rates to be lower at higher ages for adults.

This paper discusses the need to decide whether to have a national adjustment for the 2021 Census and highlights the potential ways demographic analysis might be considered for use. The aim is to gain direction on how any adjustments might be implemented and to what datasets.

Note that any decision is an a priori decision about what we would do if required. Empirical data from the Census and Administrative data will ultimately decide what we will do.

To give context the paper provides: background and a short history of the relationship between Census and mid-year estimates rebasing; why there is a requirement to adjust for undercount in the Census data; and, how demographic analysis can be used. Moving towards the 2021 Census and with the increased use of administrative data, including trying to produce statistically adjusted population estimates, it raises a larger question of whether a national adjustment is needed there and/or whether Census data is adjusted to the administrative data.

There is a need to decide whether to have a national adjustment for the 2021 Census.

There are several options considered in this paper. The key decisions are:

* Do we apply a national adjustment to the Census database if quality assurance suggests any age sex pattern may not be optimal, or consistent with what we know of migration age sex patterns, when the figures are seen?
* Do we adjust the Census database to reflect the administrative data?
* Do we adjust the administrative data to reflect demographic analysis?
* Should we consider not nationally adjusting neither dataset but allowing Population Statistics Division to produce their best estimate of population from the Census and the administrative data sources?

The following need to consider these options and feedback:

* The leads responsible for the National Statistics on population and migration. (Richard Pereira, Kerry Gadsdon, Sarah Crofts, Jay Lindop, Mike James, Becca Briggs) (January 2020)
* The leads responsible for producing Census estimates and administrative estimates. (Jen Woolford, Steve Woodland, Owen Abbott, Cal Ghee, Ann Blake, Jon Wroth-Smith) (February 2020)
* The external assurance panel for Census and Transformation (March 2020 – this will be the version distributed to the Census leads above)
* Final decision from senior staff in ONS Emma O’Rourke, Liz McKeown Iain Bell, Prof Sir Ian Diamond (April 2020).

**Background**

The current population statistics system is based on carrying out a Census once every decade. In between the census years, estimates are made of the components of change for births, deaths, internal, cross border (with Scotland and Northern Ireland) and international migration. Births and deaths are considered to be of high quality with complete coverage as they are provided through a registration system. Internal migration is estimated from administrative data and is a zero sum at the England and Wales level. Cross border migration is also estimated from administrative data and is a relatively small component for England and Wales combined (it is a relatively larger component for Scotland and Northern Ireland and an important component between England and Wales). International migration estimation is derived mainly from the International Passenger Survey, although administrative data is beginning to be integrated.

**Demographic Analysis**

Demographic analysis includes measures of the dimensions and dynamics of a population. There are some known and understood patterns, such as the shape of fertility, migration and mortality age specific curves and the fact that change in population over time can only come from births, deaths and migration. With births and deaths known implied net migration can be calculated, as well as comparisons with information from questions that measure migration.

There are some very stable measures, one key measure being the sex ratio at birth, where around 105 boys are born for every 100 girls. The sex ratio is a useful and sensitive measure for checking the veracity of population numbers. It can be calculated for a population based on births and mortality alone to provide a model sex ratio by age in the absence of migration. Any deviation must therefore be the result of differential net migration numbers for men and women. Such calculations can also be used to infer the sex ratio of the EW born who are likely to be alive but abroad.

The strength of demographic analysis means it is a valuable tool for assessing the quality of the Census and in adjusting the estimates.

**Purpose of this paper**

This paper is to highlight the potential ways demographic analysis might be considered for use and to gain direction on how any adjustments might be implemented *and to what datasets*. Options are described later in the paper. To give context the paper also provides a short history of the relationship between Census and mid-year estimates rebasing. Note that demographic analysis essentially considers plausibility, it cannot give the ‘right’ answer/number. If it could, we would not need to use the Census as the basis of population estimates, nor looking at producing demographic figures from administrative data.

**A short history of national adjustments and Census rebasing**

The 1981 Census is considered to be the last almost complete Census with a usual residence population under count of only around 0.5%[[1]](#footnote-1). This was a ‘person present’ based Census so when population estimates were made based on the Census an addition was made for people temporarily abroad and subtraction for anyone not considered usual residents. This net addition was made to the Census numbers to form the mid-year estimates and not the Census itself.[[2]](#footnote-2)

The 1991 Census was perceived to have a high level of undercount but did not have a coverage survey of sufficient size to make an adjustment. Therefore, the rolled forward estimates at a national level from 1981 were used to create the total for mid-year estimates, with the Census used the change the sub-national distribution. An examination of the differences between National MYEs and Census attributed them mainly to Census error.[[3]](#footnote-3) Again no adjustment was made to the Census data.

The 2001 Census was the first to use dual system estimation techniques to try and estimate the level of undercount in the Census. This analysis was used to populate the Census database with estimated missed people. The aim was to produce a ‘One Number Census’ so that Census was adjusted for undercount and the totals from the adjusted Census would be used in rebasing the mid-year population estimates. The total estimated figure was over a million lower than the rolled forward mid-year estimates from 1991. It was, therefore, decided in rebasing the population that the difference should be spread over two decades, with around a third of the difference in the years 1982-1991 and the remainder in the period 1991-2000. However, there were still two issues. Firstly, some local authorities, in particular Manchester and Westminster, where the DSE process had not worked sufficiently because of low response, mainly relating to issues with Census processes (address register, staffing issues), and the population was too low. Secondly, the adjusted Census figures showed a marked drop in the sex ratio at around ages 18-20 that had previously been rejected in the consideration of the 1991 Census. Using the Longitudinal Study an estimation was made of men potentially missing and alongside the adjustments made to local authorities where the Census process had not worked optimally, other local authority adjustments were made. Both these adjustments were made to the mid-year estimates and not the Census. The overall number added through both adjustments was 275 thousand. The missing men mainly being young, aged 20-40s, the adjustments for the LA issues being for all age sex groups.[[4]](#footnote-4)

When similar issues with sex ratios were seen in the 2011 Census the contingency of making a national adjustment was invoked. As in 2001, the judgement was made using evidence from the ONS Longitudinal Study, which was corroborated this time by evidence from the Lifetime Labour Market Database (L2), that the number of males was too low relative to the number of females.[[5]](#footnote-5) A national adjustment was made to the Census database with the MYEs being directly based on the resulting numbers.[[6]](#footnote-6)

**Potential use of demographic analysis**

Demographic analysis can be used to adjust data so it fits certain patterns and trends, so for example, population at older ages might be adjusted to provide more plausible mortality patterns, if we trust the age/sex deaths data from registration. Demographic analysis based on sex ratios has been used in England and Wales in both 2001 and 2011 to adjust population data to shift the sex ratio at ages in the 20s to 40s to a pattern that was closer to that found in the Longitudinal Study, given expected non response derived from previous Censuses.

**Why 2021 may be different**

While the 2021 Census is different to previous Censuses in that it is designed to be online first, the basic approach to coverage is unchanged from 2001 and 2011. Methods have been finessed and 2021 will also be different in that, with online collection, data is likely to be processed without the geographical silos of estimation areas – allowing a quicker view of the whole country as response develops. But, similar issues as in previous Censuses are likely to occur. This is because it is perceived that some groups, particularly men aged in their 20s and 30s, violate the assumptions that need to hold for dual-system estimation to work.[[7]](#footnote-7) The other difference in 2021 is that ONS has a major transformation programme to produce population statistics from administrative data.[[8]](#footnote-8)

There is an expectation that around the Census results and the administrative data (which will also be coverage adjusted using a survey approach) there will be no surprises – I understand this to mean that the data will show similar results and any differences will be explainable. Mid-year population estimates are most likely to be rebased to the Census as, based on current plans, the coverage survey sample sizes required to adjust the administrative data down to local authority level will not be available in the time scale required.

Administrative data may play into any adjustment in two ways. At the aggregate level it could be used to help calibrate an adjustment, either the national level and/or for the geographic distribution of any national adjustment. This could be instead of, or in addition to, the use of the ONS Longitudinal Study analysis which uses a Markov Chain approach calibrated to the change between 1971 and 1981, where response was much higher, to estimate a level of males relative to females. This was used in both 2001 and 2011.[[9]](#footnote-9) It is also planned to link the 2021 Census responses to the micro data that form admin-based population estimates (ABPE) prior to adjustment, which may help in measuring the quality of the Census coverage This note has been written without exact knowledge of how such linkage is planned to be used in the statistical design.

Whichever way the Census and administrative data are used and combined, there is likely to be much more information around the time the Census is being analysed.

This plethora of statistics available makes trying to work out *how* to use demographic analysis and *whether* and *where* to apply any resulting adjustments is something that requires a conscious decision and strategy. Further, each data source will have strengths and weaknesses that need to be understood, some of which will only become apparent as the results emerge.

Table 1 outlines some potential different options for using demographic analysis and relating Census data, admin-based population estimates (ABPEs) and mid-year estimates together. There are many possibilities. Demographic analysis could be used to adjust either Census or ABPEs. Census and ABPEs could be used to adjust the other. MYEs could be based on adjusted or unadjusted ABPEs or adjusted or unadjusted Census, or some combination of the two. Any demographic adjustments could either be made directly to the Census and/or ABPEs or adjustments could only be applied to the MYEs with an explanation for the disconnect.

For information Annex A illustrates some of the issues around sex ratios. It is worth noting that the sex ratios in the ABPEs are more extreme, however they are unadjusted for coverage so are highly provisional.

Table 1 High level assessment of options

| Option | Detail | Pros | Cons |
| --- | --- | --- | --- |
| 1. Adjust Census database to Admin data – and rebase MYEs to this
 | In this situation we would adjust the Census data to fit age and sex totals provided by administrative data (suitably adjusted). This would not involve any direct demographic analysis adjustment. Adjustment would be at the minimum nationally by age and sex although could be taken down to lower geographical levels. MYEs would then be based on both census and admin data as they would be the same. | * Figures are all aligned
* No need for a further national adjustment
 | * Gives primacy to the admin data devaluing Census
* Can admin data deliver figures for March 2021 in time?
* Questions why coverage survey is needed of both admin data and Census
* Likely to leave an unusual sex ratio
 |
| 1. Adjust Census database using demographic analysis (e.g. sex ratio) and rebase MYEs to this
 | Use information on sex ratios from other sources (Longitudinal Study, ABPEs, other admin sources e.g. RAPID) to adjust the Census database. Adjustment would be at the minimum nationally by age and sex although could be taken down to lower geographical levels. MYEs would then be based on the results. | * In line with what we did last time
* MYEs and Census aligned
* Uses admin data in supporting any demographic analysis used
* Admin data may be able to help with geographic distribution
 | * May make people question the admin data
* May suggest demographic analysis needed to adjust admin data
 |
| 1. Adjust Admin data to Census estimate and rebase MYEs to this
 | In this situation we would use the Census results to provide an adjusted admin-based set of results. This would not involve any direct demographic analysis adjustment. MYEs would then be based on these results. | * Figures are all aligned
* No need for further national adjustment
* Gives primacy to the Census
* Adjustment of admin data can be done later
 | * Suggests future Census needed to adjust admin data
* Could suggest admin data is weak (will depend on differences)
* Questions why coverage survey is needed of both admin data and Census
* Likely to leave an unusual sex ratio
 |
| 1. Do not adjust Census database but generate MYEs based on Census data adjusted to Administrative data
 | No adjustment made to Census but MYEs produced with population numbers derived using administrative data (adjusted ABPE). | * No national adjustment to Census so may allow census figures to be released more quickly.
* MYEs based on admin data so can be consistent going forward
 | * MYEs not aligned with Census
* May cause questions about the quality of Census
* May not be quick to adjust MYEs to Census admin data mix
* If Census is still important may still suggest future Census is needed
 |
| 1. Do not adjust Census database but generate MYEs based on Census data adjusted using demographic analysis (sex ratio)
 | No adjustment made to Census but MYEs produced with population numbers derived using demographic analysis to produce additions or subtractions to the Census base that are not included in Census. | * No national adjustment to Census so may allow Census figures to be released more quickly.
* Administrative data could be used to help with demographic adjustment of MYEs.
 | * Mostly ignores administrative data
* Administrative data and MYEs not aligned
* Census data not aligned with MYEs
 |
| 1. Don’t make any adjustment- simply rebase MYEs to Census
 | Take the Census as truth without a national adjustment | * Census data published more quickly
* Census and MYEs consistent
 | * Ignores Administrative data
* Likely to leave an unusual sex ratio
 |
| 1. Don’t make any adjustment- simply rebase MYEs to ABPE
 | Take the Adjusted ABPEs as truth without a national adjustment | * Census data published more quickly
* Administrative data and MYEs consistent
 | * Ignores Census data
* Likely to leave an unusual sex ratio
 |
| 1. Do not nationally adjust Census or Administrative data but bring together with all together MYEs to produce a new MYE using some form of modelling
 | Combining all data sources to produce a final best estimate with modelling building in best estimates by age sex and geography | * Census and admin data can be produced separately
* Neither Census nor admin data is paramount
* Some thinking along these lines has been carried out by Li Chun Zhang at Southampton University.[[10]](#footnote-10)
 | * Untested method
* Complex
* May delay rebasing MYEs to new data
* May end up in a judgement discussion similar to GDP.
 |

**Annex A The sex ratio issue**

It has previously been noted that the sex differences between male and female net migration have little effect of the sex ratio of the population.[[11]](#footnote-11) This means that the mid-year estimates effectively age forward the sex ratio found in the Census. This decade is seeing a similar issue with sex ratios in that the dramatic drop in sex ratios around ages 18-20 is not reproduced by the international and cross border migration figures recorded in the intercensal period.

Figure 1 below illustrates the issue by showing the sex ratio in the 2011 MYEs (which are derived from the 2011 Census) and the 2018 MYEs. It shows that the 2011 pattern is simply being aged forward for those aged in their 20s. Either the underlying pattern is incorrect or our measures of migration by age and sex are incorrect (or both). Note the differences at older ages relate to differential mortality between men and women.

Figure 2 shows that the dip in the sex ratio is even sharper in ABPE v3. The pattern is not aged forward so there is a consistency. However, it does therefore imply a completely different age sex pattern for migration than would be supplied by the current migration data. Not enough work has been done yet to examine the sex ratio pattern of implied net migration (or gross migration) flows from the ABPEs but it is likely not to be credible.

Figure 3 suggests that there would have to be quite large differences in estimation to make the male change over time look similar to the females in the ABPE. It compares the single age population in 2011 with their cohort age in 2016. This gives the change resulting from net migration and mortality for each cohort. For younger ages the numbers for mortality are small (as seen by the lack of effect on the natural sex ratio – figures 1 & 2). Note this is not exact as the reference date is Census day for 2011 but Mid-year for 2016 – however any ‘error’ introduced by this is not likely to affect the patterns seen. For such changes to occur it would suggest very different net migration patterns between males and females. This needs to be explored further by the teams attempting to produce migration flows from administrative data.

Figure Sex ratio Mid-year estimates 2011 and 2018 and 2011 natural sex ratio

Figure Sex ratio ABPEv3 2011 and 2016. MYE 2016 and MYE sex ratio aged forward from 2011

Figure ABPEv3 Cohort change\* in size between 2011 and 2016

\*For example, the figure for males age 20 in 2016 is the difference between the number of males age 15 in 2011 and the number of males aged 20 in 2016 – a negative number indicates the number is lower in 2016 than 2011.

1. *The 1981 Census of Population in England and Wales* A R Thatcher Population Trends no 36 pp5-9 [↑](#footnote-ref-1)
2. *Population Definitions* Population Statistics Unit OPCS Population Trends no 33 pp21-25 [↑](#footnote-ref-2)
3. How complete was the 1991 Census? (1993) Population Trends no. 71 [↑](#footnote-ref-3)
4. Local Authority population studies 2004 [https://webarchive.nationalarchives.gov.uk/20160128195726/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/pop-ests/local-authority-population-studies/index.html](https://webarchive.nationalarchives.gov.uk/20160128195726/http%3A//www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/pop-ests/local-authority-population-studies/index.html) [↑](#footnote-ref-4)
5. Office for National Statistics (2012k) Making a national adjustment to the 2011 Census. Available

at: <http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-data/2011-first-release/first-release--quality-assurance-and-methodology-papers/making-a-national-adjustment-for-residual-biases.pdf> [↑](#footnote-ref-5)
6. Note that there are some minor differences in geography as special tables are run to move some armed forces personnel for MYE definitions. [↑](#footnote-ref-6)
7. Key assumptions for DSE are Closed target population (no opportunity for people to enter or leave the population of

interest)

• independence between the two sources (in this case Census and Coverage Surveys (the likelihood of being recorded on one has no relationship with the likelihood of being recorded on the other)

• homogeneity of capture of individuals (all individuals have the same likelihood of being

captured in a list)

• no erroneous inclusions in either

• perfect linking between the two sources [↑](#footnote-ref-7)
8. Administrative Data Census Research Outputs <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs> [↑](#footnote-ref-8)
9. See section 5 in Longitudinal Study 2011 Census Linkage Report

<http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/quality/quality-assurance/longitudinal-study-2011-census-linkage-report.pdf> [↑](#footnote-ref-9)
10. Paper On provision of UK neighbourhood population statistics beyond 2021 (unpublished) provided in personal correspondence with Paul Smith [↑](#footnote-ref-10)
11. Steve Smallwood and Sofie De Broe (2009)Sex ratio patterns in population estimates Population Trends no 137 <https://link.springer.com/content/pdf/10.1057/pt.2009.33.pdf> [↑](#footnote-ref-11)