Suppression of MSOAs with low population

Suppression of Middle Super Output Areas (MSOAs) with low population is necessary to ensure robust estimates of young participation rates. Excluding MSOAs with a population of less than 50 people is the optimal choice, because it maximises the stability of the estimates of the underlying participation rates while still retaining most MSOAs for analysis.

1. Participation rates under TUNDRA (tracking underrepresentation by area) are calculated based on five years of pupil cohorts. In order to ensure that TUNDRA is based on robust estimates, it is important to assess the participation rates for areas with a low population. An area with a lower population would lead to higher fluctuations in participation rates from one year to the next, because a marginal increase or decrease of an individual in the area would have a large impact on the participation rates. In this section, we aim to investigate the level of the population where suppression should occur to maintain stable and meaningful participation rates.

2. Table 1 demonstrates the fluctuations in participation rates, which are dependent on the population size of the statistical geography. In Table 1, x represents the population size, and 1/x represents the amount of impact one additional or one less individual has on the participation rates for a population of size x. This creates the fluctuation in percentage points, where for example a population size of 10 means that one individual can increase or decrease the participation rate for that area by 10 percentage points. As the population size for the area increases, the fluctuation decreases. With a population of size 50, an individual can influence the participation rate by 2 percentage points.

Table 1. Relationship between population size and participation rate fluctuations

x	1/x	Fluctuation in percentage points
10	0.1000	10
20	0.0500	5
30	0.0330	3.3
40	0.0250	2.5
50	0.0200	2
60	0.0160	1.6
70	0.0140	1.4
80	0.0125	1.25
90	0.0111	1.11
100	0.0100	1

Why are MSOAs the chosen statistical geography?

3. MSOAs are chosen as the statistical geography for TUNDRA because of the size of its population. To create an insightful and stable measure, the size of geographical areas used should be minimised while maintaining reliable participation rates for each area. Compared to other statistical geographies of lower levels, such as LSOA (Lower Super Output Area) and OA (Output Area), MSOAs are larger areas, which contain a larger population for more stable rates.

4. Table 2 displays information regarding the size of each statistical geography based on the 2009-10 to 2013-14 Key Stage 4 year-cohort population. As can be seen, the minimum population per area for each statistical geography are all low and will yield meaningless participation rates, so some areas will be excluded no matter which geographical classification is used. The mean population also indicates that OA may not be suitable, because the mean of 16.3 people per OA area suggests a fluctuation between 5-10 percentage points as demonstrated by Table 1.

Table 2. Population information of each statistical geography

Statistical geography	Number of areas	Minimum population per area	Mean population per area	Maximum population per area
MSOA	6791	6	405.8	1158
LSOA	32828	1	83.9	277
OA	168736	1	16.3	167

5. Figures 1a, 1b, and 1c show the frequency of the different population sizes for each statistical geography in the order of MSOA, LSOA, and OA. A population suppression cut-off size of 50 has a minimal impact on MSOA, some impact on LSOA, and suppresses most OAs. This makes OA and LSOA unfit for the purpose of TUNDRA due to the exclusion of a large number of geographical areas.

Figure 1a: Population size of MSOAs

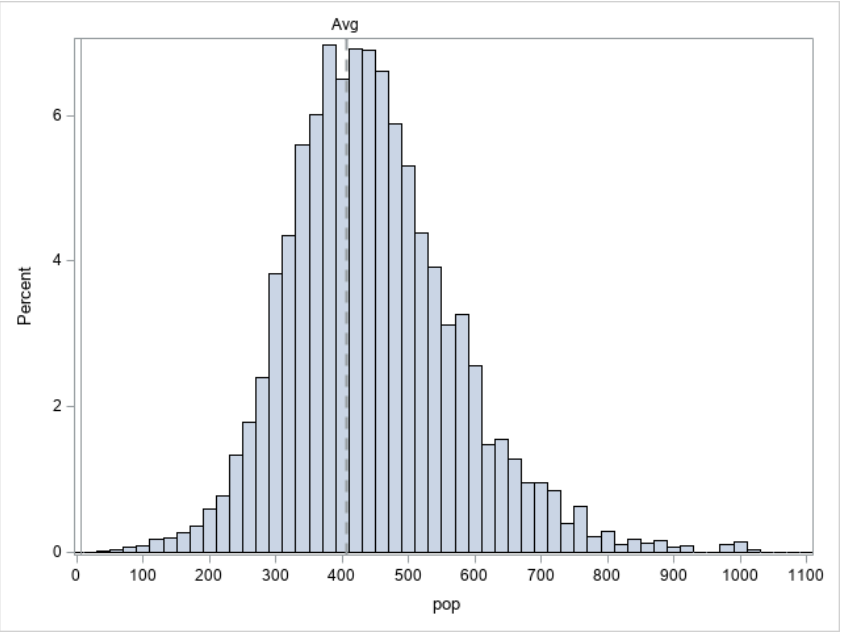


Figure 1b: Population size of LSOAs

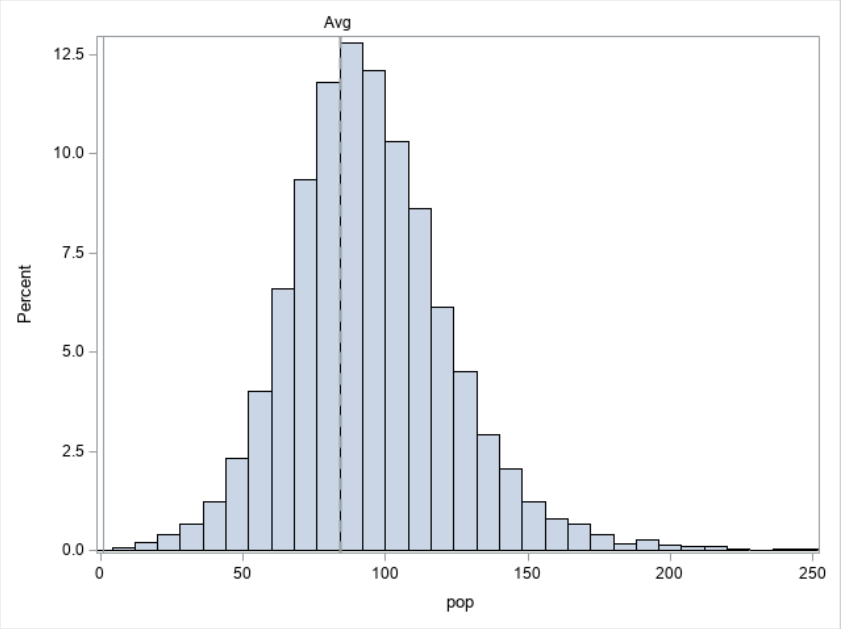
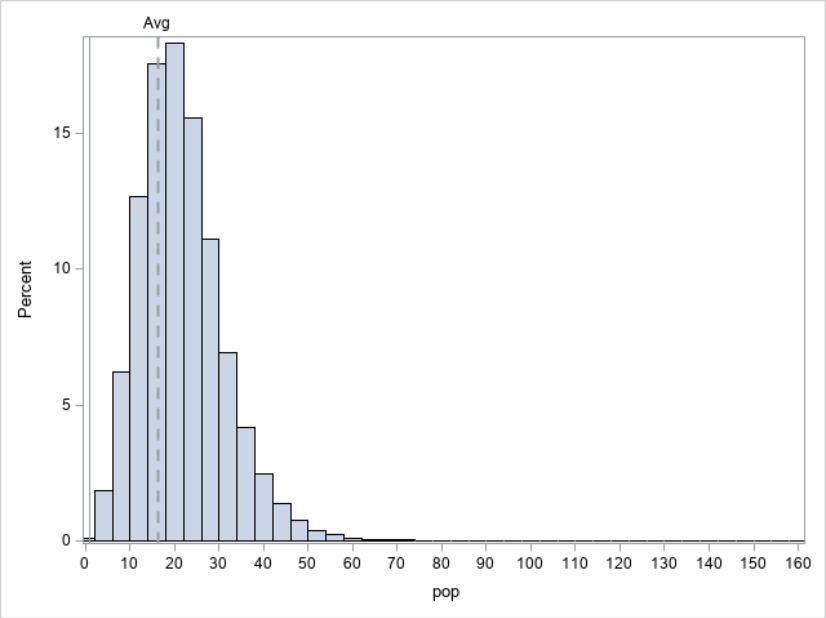


Figure 1c: Population sizes of OAs



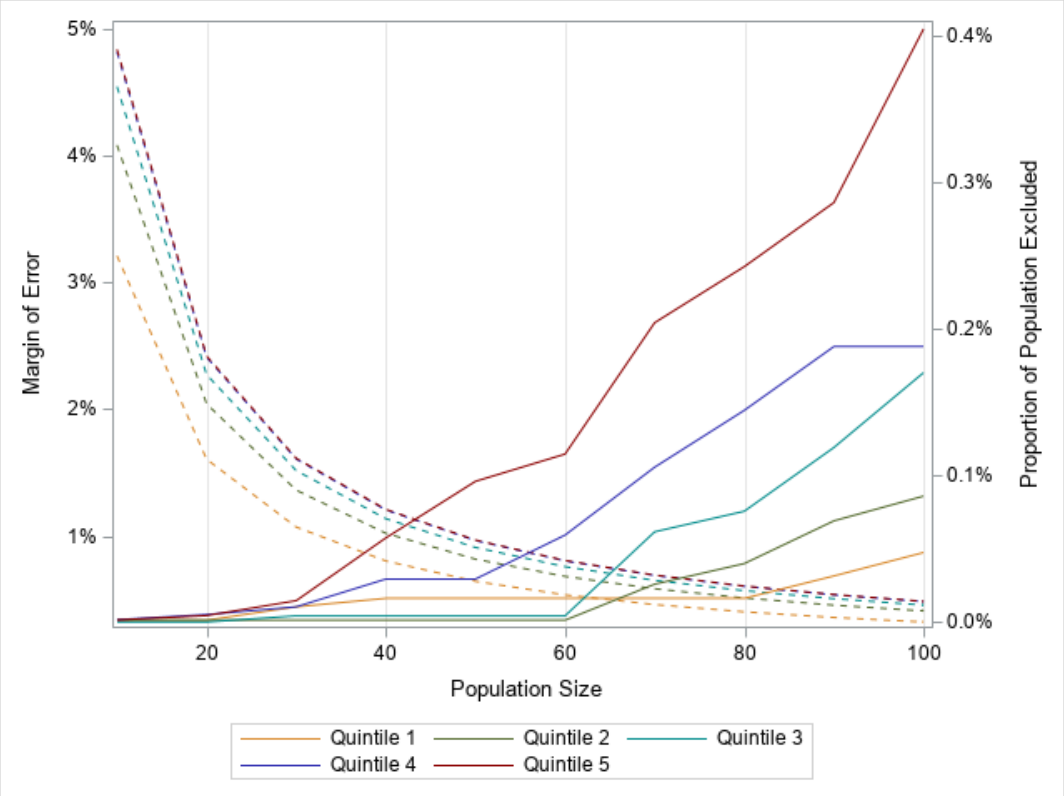
Selecting the size of population to suppress: impact on different quintiles

6. Having selected MSOAs, the impacts of suppression on the rates for each MSOA and the quintiles derived from them and how the impacts vary depending on the size of the population suppressed can be examined.

7. Using the mean of each quintile, the confidence interval and margin of error on the MSOA participation rate estimate is calculated as the size of the MSOA population is varied. The confidence level chosen for this investigation is 95 per cent. For small population sizes, we can expect the potential margin of error to be large, meaning that we are less certain that the true participation rate of the area is captured in our calculated participation rate due to the population size.

8. Figure 2 displays the trade-off between margin of error and the proportion of population excluded as the population size changes. It is important to note that the intersection between these curves for each quintile does not represent the optimal or most efficient population suppression size. The intersections only occur as a result of the scales used in the different y-axes. Upper quintiles are more impacted in the trade-off – for example, a higher proportion of people in quintile 5 are excluded as a result of increased population suppression size compared to quintile 1. The higher the population size is, the lower the margin of error, and the lower the margin of error the more stable the participation rates are. However, this comes at a cost of excluding the population. Suppressing certain MSOAs with low populations would mean a certain proportion of the population are not included in the calculations of participation rates, which would mean it is less applicable to England as a whole.

Figure 2. Trade-off between margin of error and proportion of population excluded, by quintile



Note: dash lines represent margin of error, solid lines represent proportion of population excluded

9. This relationship is explored further in Figure 3a, which plots the margin of error against the proportion of the population excluded directly, linked together through the population suppression size. Figure 3a clearly shows the convex nature of the relationship, where at high levels of margin of error the proportion of the population excluded does not change by a large amount, and at minimal values of margin of error, the proportion of the population excluded is impacted. The figure confirms the previous assumption that suppressing areas at levels of population below 50 (ex. 40, 30, 20) are not effective at obtaining accurate participation rates, because the margin of error would still be high for these cases. In relation to Table 1, at population sizes below 50, an individual participating or not participating in higher education could increase or decrease the participation rates by more than 2 percentage points.

10. Figure 3b focuses on the population sizes between 50 to 100. There is a clear difference in the magnitude of the relationship between different quintiles. The population of quintile 1 is relatively inelastic and is not impacted much by changes in population size or margins of error. It hovers at approximately 0.04 per cent at each population suppression size. The population of quintile 5 is a lot more elastic when it comes to changes in the population size used for suppression. At a population size of 50, the proportion of population excluded is under 0.1 per cent. This rises quite dramatically as the population size increase to over 0.4 per cent when the population size is 100. This is quadruple the number of people who are excluded when the population size of 50 is used. The fact that upper quintiles are impacted more is due to upper quintiles being assigned to smaller MSOAs, so a suppression of these would mean they are excluded.

Figure 3a. Relationship between margin of error and proportion of population excluded, by quintile

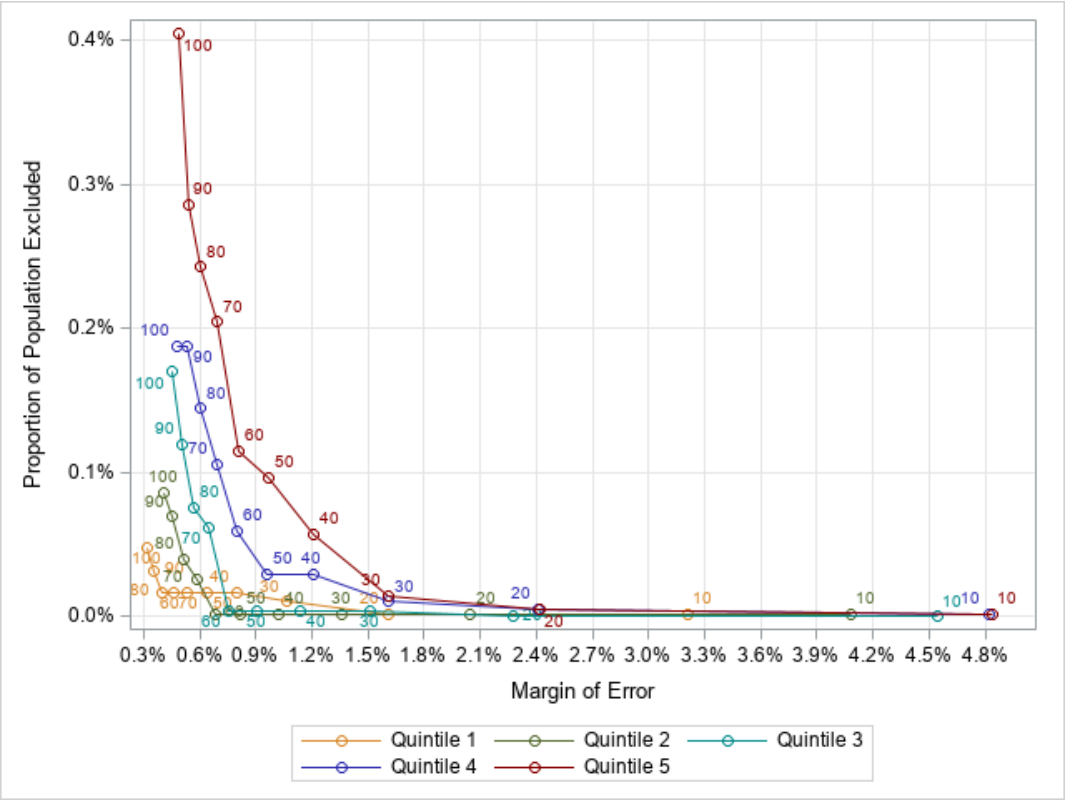
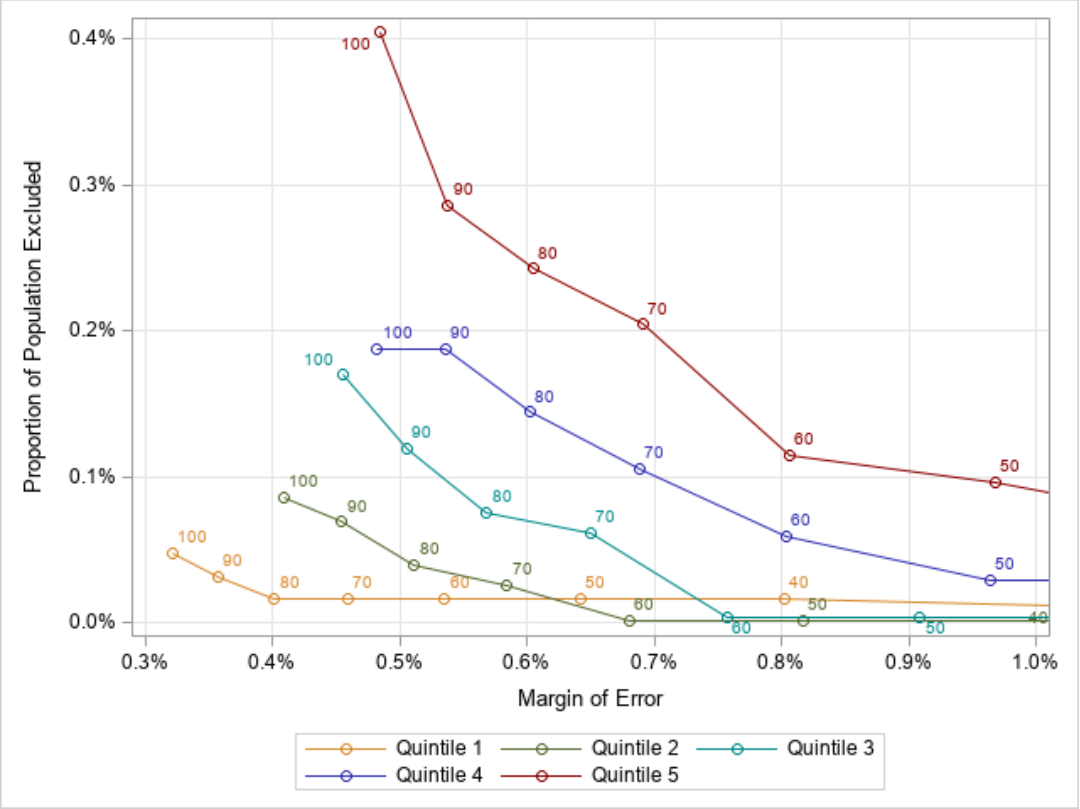


Figure 3b. Relationship between margin of error and proportion of population excluded, by quintile (where population below 50 is displayed)



11. Choosing an appropriate level of population to suppress would be difficult, since what is considered an acceptable proportion of excluded population and acceptable margin of error is subjective. However, because the trade-off is different for each quintile group, it plays a pivotal role in the decision-making process. In order to not disproportionately exclude people from upper quintiles, the population size with the lowest disproportionate impact should be chosen. As previously stated, because suppressing population sizes of less than 50 is ineffective, 50 is the level which optimises the suppression process. If MSOAs with less than 50 people are not included in the analysis for TUNDRA, it would minimise the margin of error, maximise the proportion of the population included in the analysis, and reduce bias between higher and lower quintiles. Using a population of 50 as the cut-off point for suppression results in a margin of error of less than 0.1 per cent for each quintile, and excludes less than 0.1 per cent of the population for each quintile. A total of 27 MSOAs are suppressed, and so TUNDRA quintiles will not be created for these suppressed MSOAs.