

Invitation to comment: Experimental release of the TUNDRA participation classification based on LSOAs

Overview

This document describes a new version of the area-based measure called TUNDRA (tracking underrepresentation by area), which uses tracking of pupils in state-funded mainstream schools in England to calculate young participation. It groups areas across England into five groups, called quintiles, based on the proportion of young people from that area who participate in higher education at age 18 or 19 years. The Office for Students (OfS) has previously released the experimental higher education participation classification TUNDRA based on the geographic areas called Middle Layer Super-Output Areas (MSOAs). This is described on the OfS website.¹

The aim of this invitation to comment is to gather information and feedback concerning the potential release of a TUNDRA classification based on the smaller Lower Layer Super-Output Area (LSOA) geography in England.

For clarity, the experimental TUNDRA classification based on MSOAs is now referred to as TUNDRA MSOA, and the experimental TUNDRA classification based on LSOAs is now referred to as TUNDRA LSOA. The common methodology is simply referred to as TUNDRA.

There are three documents in this release. The other two are technical, detailing specific supporting analysis. This document is a discussion of implications using the new TUNDRA LSOA classification, and questions for which we would very much like feedback.

Background information

The aim of a participation classification is to understand how participation in higher education varies geographically across the country, in order to help identify areas of persistent low (or high) participation in higher education. The OfS has previously generated the POLAR (participation of local areas) classifications for this purpose. The experimental TUNDRA methodology takes advantage of new technologies to create a participation classification. A full description of the POLAR and TUNDRA methodologies is available on the OfS website.²

There are many ways in which the UK may be split into smaller areas when reporting statistics. One of these is based on census information and is created by the Office for National Statistics (ONS). Small areas (called Output Areas, or OAs) are grouped into Lower Layer Super-Output Areas (LSOAs), which are in turn grouped into Middle Layer Super-Output Areas (MSOAs). Detail of the Output Area geographies is available on the ONS website.³

¹ See www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/

² See www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-polar/

³ See www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography#output-area-oa

The OfS uses MSOAs for the POLAR4 classification and the first experimental TUNDRA MSOA classification. Both POLAR4 and the initial experimental TUNDRA MSOA classification each give one participation quintile for a whole MSOA.

The LSOAs, which are smaller than the MSOAs, are used nationally for statistics such as the Index of Multiple Deprivation (IMD). There are 32,844 LSOAs in England, which is almost five times the number of MSOAs. A classification based on LSOAs would give a participation quintile for each LSOA.

Updating the classification to take advantage of the smaller LSOAs may be useful to inform understanding of local participation patterns, enabling identification of small pockets of low or high participation. However, the smaller population of students in each LSOA could result in a less robust classification. We would like feedback to understand if the increased detail would be sufficiently useful to outweigh any adverse consequences.

The TUNDRA classification

The experimental TUNDRA MSOA classification is based on the population of 16-year-old students who live in England and attend state-funded mainstream schools in England, for the academic years between 2010 and 2014. We track these students to identify whether they appear in higher education either two or three years later (which will be at age 18 or 19, as this is the definition of young participation). We calculate a participation rate for the students living in each area – this was the MSOA in the original experimental TUNDRA MSOA release, but is now the LSOA for the classification considered here. The areas are then ranked by their participation rate and grouped into quintiles. Areas in quintile one have the lowest participation in higher education, and areas in quintile five have the highest participation in higher education. The full TUNDRA methodology is available on the OfS website.⁴

The robustness of a TUNDRA classification generally increases as the number of students in each area increases. One way to have a larger population is use of a larger physical area, as larger physical areas will tend to have a larger number of students living in them. There are two further methods which TUNDRA uses to help ensure a classification is robust. First, it combines students from five successive years of GCSEs. For example, the experimental TUNDRA MSOA classification tracks students who are 16 in any of the years 2009-10, 2010-11, 2011-12, 2012-13, and 2013-14. However, even after combining five years of GCSE students there are still some areas which have very few students. For the MSOA-based classification, if there are fewer than 50 students from an area, the data is suppressed (not published) for that area as it may be unreliable.

LSOAs are smaller than MSOAs and are nested in MSOAs: on average an MSOA consists of four or five LSOAs. As well as being smaller geographically, the number of students living in an LSOA will be less than the number of students living in the MSOA it sits in, and typically it will be considerably smaller. Even after combining five cohorts of students taking GCSEs, more than 10 per cent of the LSOAs have fewer than 50 students. Publishing a classification in which more than 10 per cent of areas are not given a quintile, is not considered acceptable.

In order to publish the quintiles for more LSOAs there are two possibilities. First, it is possible to combine more cohorts of GCSE students. For example, increasing the number of GCSE cohorts to

⁴ See www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/

ten years would increase the number of areas with at least 50 students to track. But combining a larger number of cohorts would mean that a local change in participation would take a long time to be reflected in the (combined) local participation rate, and therefore a long time to be reflected in the quintile assigned to an area. There would be a large lag in publication.

The second option is to publish data for areas where there are fewer than 50 students in the GCSE five-year cohort. The technical analysis⁵ examines this option, and quantifies the robustness of a TUNDRA LSOA classification which suppresses publication if there are fewer than 30 students from an area. Just under 4 per cent (one in 25) of the LSOAs have a student population of less than 30 over the combined five years of data, and therefore would not have a quintile published in the classification.

How does the TUNDRA LSOA classification differ to the MSOA one?

Each MSOA is created by grouping together a small number of LSOAs, so that each LSOA fits neatly into exactly one MSOA. By comparing the quintile of the LSOA and the quintile of the MSOA it fits into, it is possible to see if the more local information of the LSOA classification differs to the information about the MSOA it is in.

A very small difference in participation (perhaps due to a single student), could cause an area to move from one quintile to an adjacent quintile. A difference of at least two quintiles is therefore needed in order to be considered important, and areas which show a difference of a single quintile are grouped with the areas which show no change in quintile and considered as not significant.

The comparison is summarised in Table 1.

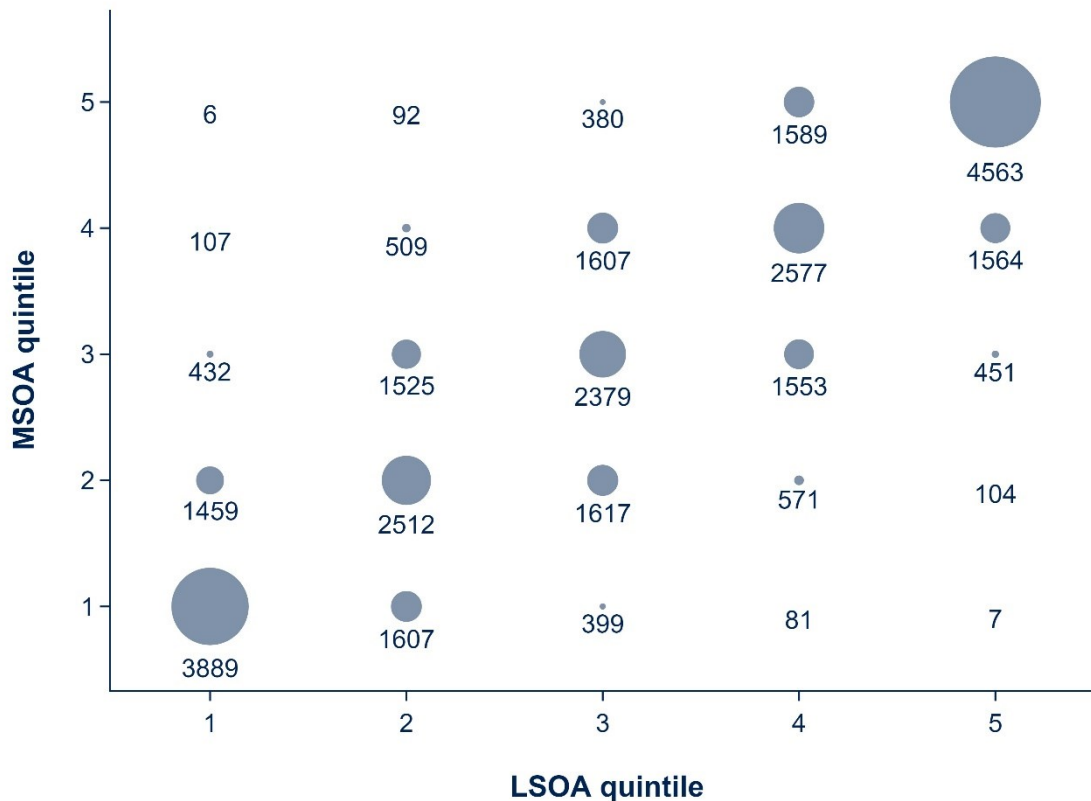
Table 1: Summary of the difference in quintile between the TUNDRA LSOA and TUNDRA MSOA classifications

Difference between LSOA and MSOA quintile	Number of LSOAs	Percentage of LSOAs
The same quintile, or one quintile difference	28,441	86.6
At least two quintiles difference	3,139	9.6
Undefined, in either the MSOA or LSOA classification or both	1,264	3.9

The differences are shown in Figure 1 (the undefined MSOAs and LSOAs are not included in this).

⁵ Available at www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/

Figure 1: Bubble plot summarising the differences between the TUNDRA LSOA and TUNDRA MSOA (undefined areas are not shown)



Looking at the very top row, this shows the areas which are in quintile five in the MSOA classification (so have the highest MSOA participation rates). Looking at the top row of bubbles from right to left, in the LSOA classification:

- 4,563 of these stay in quintile five in the LSOA classification
- 1,589 move to quintile four
- 380 move to quintile three
- 92 move to quintile two
- six move to quintile one.

The six areas which move from quintile five in the MSOA classification to quintile one in the LSOA classification are a very small proportion of all LSOAs. But this does suggest pockets of very low participation tucked into MSOAs which have very high participation. Reading across the bottom row of the bubble plot shows there are seven areas in quintile one in the MSOA classification, which are in quintile five in the LSOA classification. This identifies pockets of very high LSOA participation tucked into MSOAs with very low participation. These are the two very extremes of difference and do not represent the average case.

The majority of LSOAs (86.6 per cent) are similar to the MSOA they are contained within, being either in the same quintile or differing by just one quintile. But almost 10 per cent (one in 10) of LSOAs show an important difference from the MSOA they fit into, with a difference of at least two quintiles between the MSOA and LSOA classification.

Very roughly, each MSOA contains about five LSOAs. Selecting two MSOAs at random, these will comprise about 10 LSOAs. As one in 10 of the LSOAs differ by the parent MSOA by at least two quintiles, on average we would expect one of these 10 LSOAs to show a large difference to the MSOA it fits into. So out of the two MSOAs which are selected at random, one of them is likely to contain an LSOA which is a pocket of low or high participation in comparison.

Using the TUNDRA LSOA classification

As soon as a new TUNDRA LSOA classification is available, immediately after the classification is published, the most recent data included in the classification will be at least 18 months out of date. This happens because it is the earliest possible date which the OfS can receive the data. At this time, using the classification to inform outreach starting the following academic year would inform the outreach two years after the latest data in the classification.

A classification is based on past data, and there is the assumption that future participation from an area will be well-predicted by past participation. If there is no underlying change in the participation rate in an area, the quintile should be a good prediction for future participation.

Conversely, if there is actually a true change in the underlying participation rate in an area, it is desirable that this is reflected in the classification. An ideal classification would remain the same if there is no change in the underlying participation rate, whilst being sensitive to, and changing if there is, an actual change in the underlying participation rate. In practice, changes in the quintile of an area over time will in part be due to reasonable natural random variation in participation where the underlying rate does not change, and part due to an underlying change in the participation rate. In the stability and sensitivity analysis, reported in the technical documentation, changes to the quintile of an area over time are apportioned to either random variation where the underlying participation rate does not change, or to a true underlying change in the participation rate.

For the earliest academic year in which the classification can be used, the data used to calculate the quintiles for the LSOAs is from at least two years previously. About one in 80 of the LSOAs will already have moved by at least two quintiles compared to the classification if the latest data were to be available, so using the published classification could be misleading for this LSOA. Just over half the time, the large change in quintile will be due to an underlying change in the participation rate. For the rest of the time, it will be due to natural random variation around the true underlying participation rate. It is not possible to tell at this time which areas will have changed in the latest data until that classification is actually available. Even when the later data becomes available, it is not possible to say which areas have moved for each of the two reasons.

A year further on, the data used for the classification would be from at least three years previously. This longer delay allows for greater changes to underlying participation, and a larger random part to the participation rate (as a larger part of the tracked population will have changed). At this time, one in 30 of the LSOAs will have moved by at least two quintiles compared to the latest data and classification (if it were to be available), so using the published classification could be misleading for this LSOAs. About three quarters of the time this large change in quintile would be due to an

underlying change in participation. The rest of the time, it will be due to natural random variation around the true underlying participation rate. It is not possible to tell at this time which areas will have changed in the latest data until that classification is actually available. Even when the later data becomes available, it is not possible to say which areas have moved for each of the two reasons.

Feedback requested about the LSOA classification

We would like feedback to understand if the TUNDRA LSOA classification, with suppression below a base population of 30, would be useful. It would be really helpful if you could consider the following points and provide feedback via our short survey.⁶

If the OfS publish an LSOA classification:

- Would you use a TUNDRA LSOA classification? What would be your main use?
- Would you also use a TUNDRA MSOA classification? How would you use it alongside the TUNDRA LSOA classification?
- At the earliest time a TUNDRA LSOA classification can be used, about one in 80 of the LSOAs will have already changed by at least two quintiles and be out of date. Do you feel this is too high for your uses?
- Are there any LSOAs which are suppressed in the TUNDRA LSOA classification, which are a particular concern for your purposes?

⁶ Survey available via www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/