Public Health England

Protecting and improving the nation's health

Technical Guide

Aggregations

Contents

Introduction

When data are not available for a required area, values can be calculated or estimated from other areas where data are available. For example, rates of hospital admissions for clinical commissioning groups (CCGs), the required areas, can be estimated from local authority data, the source areas. This guidance outlines three methods for reattributing data in this way and defines the circumstances in which each should be used.

In the simplest situations it is possible to use direct aggregation techniques to calculate statistics precisely for other areas or organisations (method 1). In more complex situations, estimation techniques must be used. Estimation methods apply weightings to each source area, based on the proportion of the population in the source area that can be attributed to each required area, and then aggregate the weighted data by the required areas. Where possible these estimates are calculated using the numerator and denominator breakdowns required to produce the statistic for the source areas (method 2), but where only the final statistic is available for the source areas this can be used directly (method 3).

Glossary of terms

There can be much confusion and ambiguity in describing the various scenarios for the reattribution methods being explained, so for the purposes of this guidance the following terms and definitions apply:

Area: for ease of reading, the document generally refers to *areas* – these can be geographical areas, such as local authorities or CCGs, but they can also be population breakdowns by organisational responsibility, such as general practice (GP) or CCG registered populations. The term *geographical areas* is used when specifying geographically defined populations only, for example when describing estimating between resident and registered-based populations.

Coterminous/ hierarchical: describes the relationship between source areas and required areas when it is such that required areas can be built precisely by grouping the source areas together without splitting any of the source areas and without leaving any parts of the required areas unaccounted for.

Whilst *coterminous* is used when referring to resident populations within geographical areas and *hierarchical* is generally used when referring to registered populations for which organisations are responsible, it is not necessary to distinguish between the two for the purpose of applying the methods described in this document. What is important is to understand which specific population groups are represented in each of the source areas and required areas.

Examples of coterminous/hierarchical relationships include:

- grouping geographical resident populations from lower tier local authorities (LTLAs) to regions
- grouping organisational registered populations into higher organisations, such as GPs to CCGs

Examples of non-coterminous/non-hierarchical relationships include:

- mapping geographical areas from wards to LTLAs
- mapping any registered population breakdowns to resident area populations (and vice versa)

- mapping registered population breakdowns for one organisation type to another at a different point in time if the relationship between the two organisation types has changed.

Aggregation methods

Building blocks of geographies and organisations

The building blocks referred to in this section can flexibly represent a variety of geographical areas or registered populations.

In the simplest scenario, data can be aggregated by grouping counts and summing them (method 1). For example, in figure 1 all the counts for each block in groups 1 and 2 can be summed to give the overall count for each group: 100 and 200 respectively. As all the counts are available for each of these smaller blocks, the counts for different groupings can also be calculated by adding them up: for example, the overall counts for groups 3 and 4 are 184 and 116 respectively.

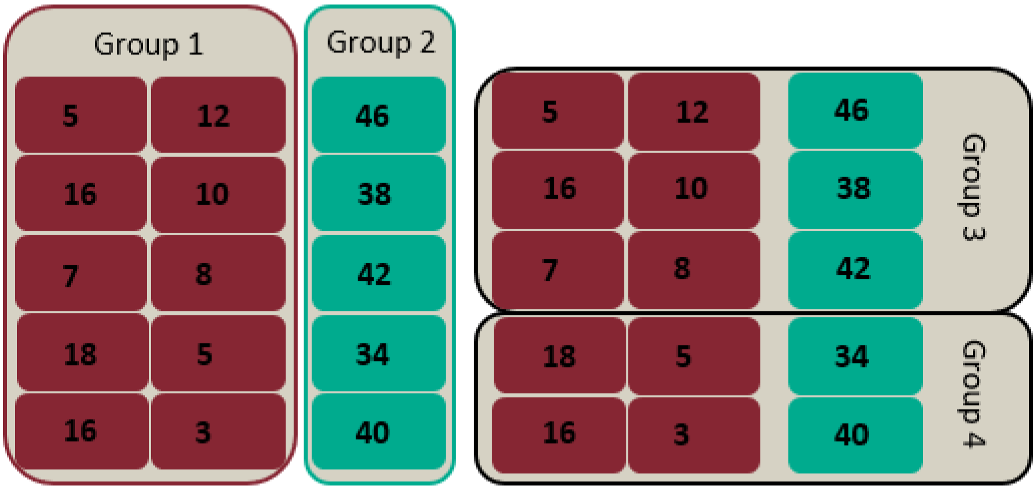

Figure 1. Counts for blocks grouped by different groupings.

When only the counts for the whole groups are known, and not the individual blocks within the groups, it is not possible to re-group the blocks into different groupings. In the top image of figure 2, the total counts for group 1 and 2, the source groups, are known, but without knowing the counts of the blocks it is not possible to sum the blocks to get the counts for groups 3 and 4, the required groups.

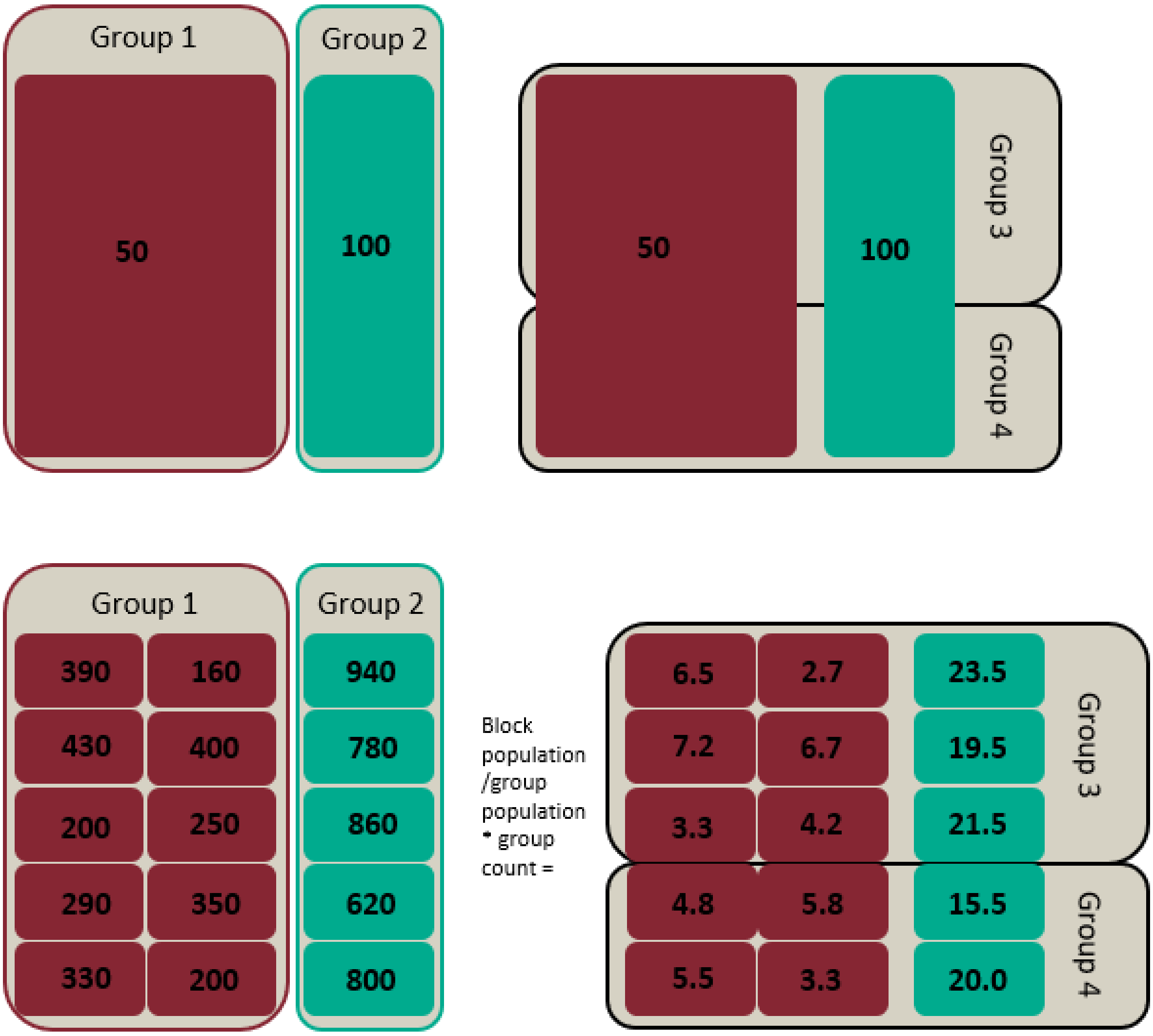

Figure 2. Top: Illustration that the counts of group 1 and 2 cannot be used to work out the counts of group 3 and 4 without knowing the breakdown on counts by smaller segments. Bottom left: populations for each block. Bottom right: estimation of each block's count from the proportion it makes up of the overall group population.

However, it's possible to estimate the counts for each of the blocks within groups 1 and 2 based on the proportion of the groups each block makes up (method 2). These estimated counts for each block can then be summed to give estimated counts for the required groups. The population size of each block is used to calculate the proportion of the group that it makes up. The bottom left image of figure 2 shows the populations for each of the blocks. Group 1 has a total population of 3,000 and group 2 a total population of 4,000. The proportion each block makes up of the total group can be calculated by dividing the blocks population by the group's total population. This proportion can be multiplied by the group's total count to give a weighted estimate of the block's count, as shown in the bottom right image of figure 2. Estimated counts for groups 3 and 4 can now be calculated by summing up the blocks' estimates, which give 95.1 and 54.9 respectively. This method assumes that the count of the larger group can be proportioned out to each of the contributing blocks based only on the population size of each group, and that there are no other factors that would cause counts to vary between each of the blocks.

This approach to estimating higher level groupings from available data applies to geographies with physical boundaries, such as local authority resident populations, and organisational breakdowns without physical boundaries, such as sustainability and transformation partnerships' (STPs) registered populations.

Low level geographies, such as lower super output areas (LSOAs) and output areas (OAs), can be used as the building blocks of higher level geographies with physical boundaries. Publicly available lookups and populations can be used to calculate the populations that live within different areas. Figure 3 shows the overlap of two CCGs residential boundaries and local authorities in Nottingham and Nottinghamshire. Nottingham City CCG and Nottingham City unitary authority are coterminous, but Nottingham North and East CCG overlaps with three different LTLAs, not encompassing any of them completely. Using LSOA populations and lookups, the total populations of the LTLAs, CCGs and the segments of areas that fall within each can be estimated, as shown in figure 3.

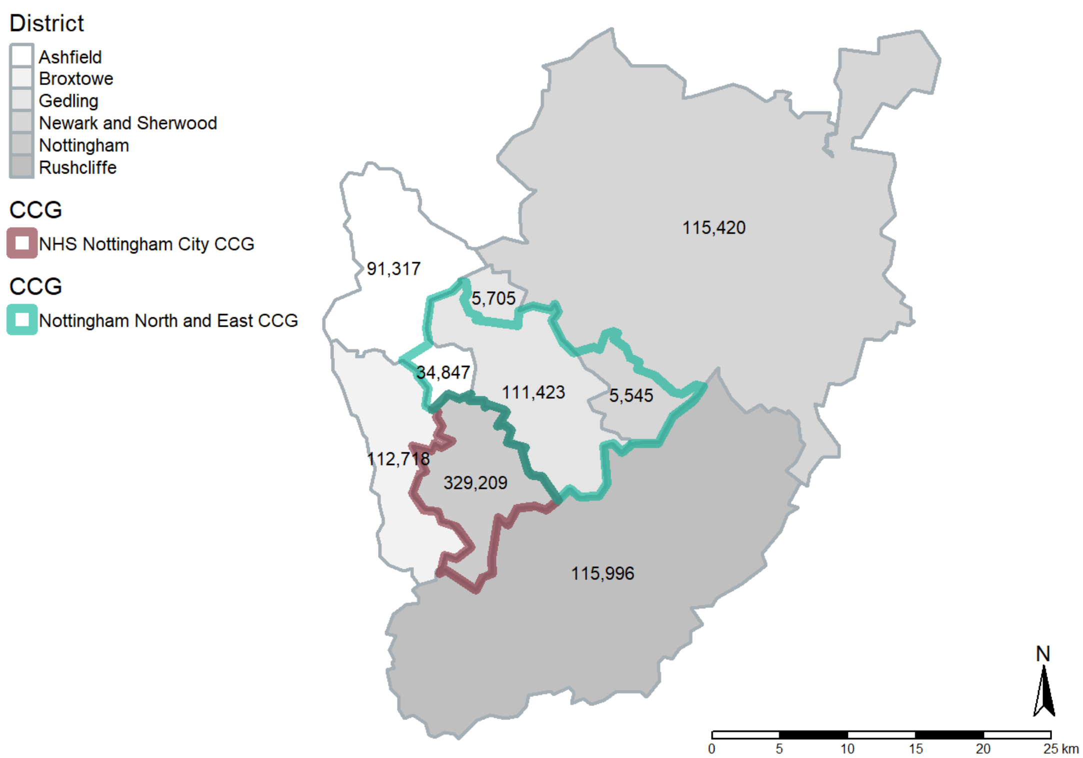

Figure 3. Map of Nottingham City CCG and Nottingham North and East CCG with overlapping lower tier local authorities. Numbers indicate 2017 populations for persons of all ages in overlapping constituent segments.

From these figures, the proportion of each LTLA's resident population that also resides in each CCG can be calculated (table 1). This gives the equivalent proportional weightings as calculated in the example above. Where data are available for LTLAs, these proportions can be used to estimate the equivalent values for the segments and then summed by the CCGs to give estimates for the CCGs.

Table 1. Proportion of lower tier local authority populations in each of the two CCGs, 2017 populations for persons of all ages

CCG	CCG population	LTLA	LA population	LA population in CCG	% LA population in CCG
Nottingham City	329,209	Nottingham	329,209	329,209	100%
Nottingham North & East	151,815	Ashfield	126,164	34,847	28%
Nottingham North & East	151,815	Newark and Sherwood	120,965	5,545	5%
Nottingham North & East	151,815	Gedling	117,128	111,423	95%

The same principle applies to organisational data. Registered populations of GPs can be used as the building blocks for higher level organisations that they belong to, such as their CCGs and STPs. In the same way, publicly available lookups and population data for GPs can be used to calculate the populations that are registered to the higher-level organisations. NHS digital publish GP registered populations broken down by patients' LSOA of residence. This gives a way of calculating the proportion of a GP's registered population that resides in any given LSOA. Using GP and LSOA lookups, these populations can be aggregated by different organisations that are hierarchical with GPs and different geographical boundaries that are coterminous with LSOAs. For example, table 2 shows the proportion of patients registered to a GP in Nottingham City CCG who reside in a Nottinghamshire county district or Nottingham City.

Table 2. Proportion of Nottingham City CCG registered population live in each Nottinghamshire county district or Nottingham City, 2018/19 GP registered populations for persons of all ages

CCG	CCG registered population*	LTLA	CCG registered population residing in LA	% CCG registered population in LA
Nottingham City	385,998	Nottingham	346,300	89.7%
Nottingham City	385,998	Gedling	19,602	5.0%
Nottingham City	385,998	Broxtowe	15,352	4.0%
Nottingham City	385,998	Rushcliffe	2,802	0.7%
Nottingham City	385,998	Ashfield	1,295	0.3%

*LTLA populations here are estimated from GP registered populations by LSOA of residence. GP populations are known to be inflated due to patients remaining on registers after they have moved, resulting in an overestimation of the LAs population compared with the ONS mid-year estimates.

Where data are available for registered populations of these CCGs, the equivalent values can be estimated for the resident populations of the local authorities using these proportions and the method outlined above. GP registered populations tend to be spread out much wider than their contractual boundaries (figure 4 left), and resident populations tend to be registered across a number of GPs (figure 4 right). This means that when calculating these proportions there will be many local authorities, or other geographical areas, where a small number of a GP's registered patients live. It also means that rates and prevalence's from registered populations cannot be used as a direct indicator of the equivalent rate and prevalence of their contractual boundaries, as the populations are not equivalent.

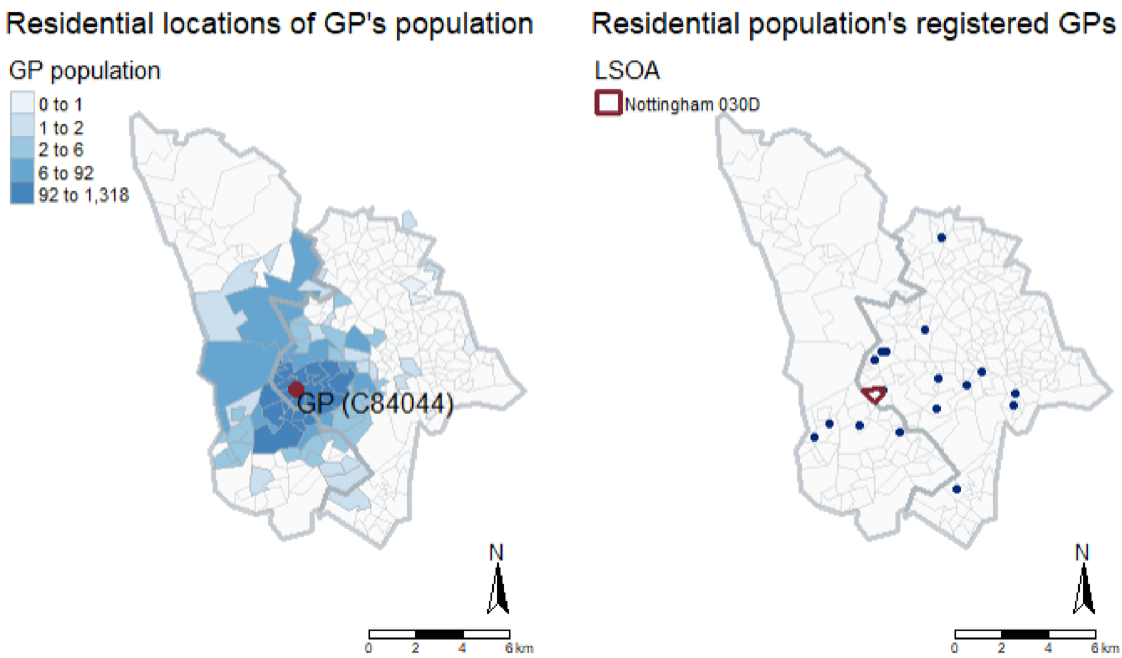

Figure 4. Left: map of the number of patients registered to the GP C84044 that live in each LSOA in Nottingham City. Right: map of GPs that residents of the LSOA Nottingham 030D are registered to.

Available data

Where all the necessary data to calculate a statistic are available for source areas, the methods outlined above can be used to calculate (method 1) or estimate (method 2) the equivalent data for the required areas from the source area and then calculate the statistic. For example, if all the numerators and denominators of a crude rate are available for the source areas, then numerators and denominators can be calculated or estimated for the required areas. These numerators and denominators can be used to calculate a crude rate for the required areas.

Conversely, if only the final statistic is available for the source areas without the numerators and denominators necessary to calculate it, the statistic can be estimated directly from the source area statistics (method 3). This is likely to be more applicable to statistics such as directly standardised rates (DSRs) and life expectancies, where data broken down by specific age bands are also required to calculate the statistic. Estimating the statistic directly like this requires weightings to be applied in a different way from that outlined above.

Estimating confidence intervals

Where the statistical values are estimated (methods 2 and 3), rather than calculated precisely (method 1) it is not appropriate to provide confidence intervals using standard methods for the statistic. This is because standard methods for calculating confidence intervals do not account for the uncertainty arising from the assumption that all small areas within a large area share the same indicator value.

Disclosure control

All of the methods described in this guidance use publicly available populations and lookup data. It may therefore be possible to back-calculate the source data from the aggregates/estimates calculated using these methods. As a result, when publishing aggregated/estimated data, the required disclosure control rules should be applied to the source data before aggregating/estimating data for other areas.

Selecting the appropriate method

In determining the appropriate method to apply, and how to apply it, it is important to understand the available data and the requirements for the reattribution with respect to the following:

1. The specific populations represented by the source and required areas and the relationship between these populations – specifically whether they are coterminous/hierarchical or not
2. Whether the source data are only available as calculated statistics or whether the data required to calculate those statistics are also available for the source populations

Table 3 provides a summary of the three methods and indicates the circumstances in which each should be applied based on the above criteria.

Table 3. Selecting an appropriate reattribution method

	Data to calculate statistic available	**Statistic available without data to calculate it**
Coterminous / hierarchical relationship between source and required areas	METHOD 1 Aggregate from source area counts and denominators	METHOD 3 Estimate directly from statistic
Not coterminous / hierarchical relationship between source and required areas	METHOD 2 Estimate numerator and denominator to calculate estimated statistic	METHOD 3 Estimate directly from statistic

Method 1: Aggregating from source data

When to apply it

This method is appropriate when the source and required areas are coterminous/hierarchical and the counts and denominators required to calculate the statistic for the source areas are available. For more complex statistics such as DSRs, this means the age-breakdown counts and denominators required to calculate the source area statistics must be available.

How to apply it

1. Aggregate (sum) the source area counts and denominators for the required areas.
2. Calculate the required area statistics in the normal way using the aggregated counts and denominators for the required areas.
3. As this method provides an accurate statistical calculation, not an estimate, calculate confidence intervals for the statistic using the standard method for that statistic.

Examples

1. The 5-year breakdown counts and denominators for all local authorities in a region can be summed to give the 5-year breakdown counts and denominators for that region. A DSR with confidence intervals can then be calculated from these regional counts and denominators.
2. The counts and denominators for each GP practice's registered population can be summed to give counts and denominators at CCG or STP level. Proportion statistics with confidence intervals can then be calculated from the CCG or STP level counts and denominators.

Notes

Lower-level area types, such as LSOAs, are coterminous with a larger range of area types (table 4). This means data available at lower levels are more flexible for aggregating up to a wider range of higher level geographies.

Table 4. Mappings of coterminous geographies

Required areas

Source areas	OA 2011	LSOA 2011	MSOA 2011	Wards 2017	Wards 2016	PCON 2010	EER 2009	LTLA 2019	UTLA 2019	CAUTH 2017	Region 2009	Country 2009	CCG 2018	CCG 2017	CCG 2016	CCG 2015	NHSRLO 2018	NHSRLO 2017	NHSER 2018	NHSER 20017	STP 2018	STP 2017	PHE Centre 2015	PHE Region 2016	CANREG 2009	CALNCV 2017	CALNCV2016	SCN 2019
OA 2011	NA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
LSOA 2011	✗	NA	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MSOA 2011	✗	✗	NA	✗	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✓
Wards 2017	✗	✗	✗	NA	✓	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓	✗	✗	✗	✓
Wards 2016	✗	✗	✗	✓	NA	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓	✗	✗	✗	✓
PCON 2010	✗	✗	✗	✗	✗	NA	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗
EER 2009	✗	✗	✗	✗	✗	✗	NA	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
LTLA 2019	✗	✗	✗	✗	✗	✗	✓	NA	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✓
UTLA 2019	✗	✗	✗	✗	✗	✗	✓	✗	NA	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗
CAUTH 2017	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Region 2009	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	NA	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Country 2009	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
CCG 2018	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	NA	✗	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗
CCG 2017	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	NA	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✗
CCG 2016	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	NA	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✓	✗
CCG 2018	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	NA	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
NHSRLO 2018	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	NA	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
NHSRLO 2017	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	NA	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗
NHSER 2018	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	NA	✓	✗	✗	✗	✗	✗	✗	✗	✗
NHSER 2017	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	NA	✗	✗	✗	✗	✗	✗	✗	✗
STP 2018	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓	✓	NA	✗	✗	✗	✗	✗	✗	✗
STP 2017	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓	✓	✓	NA	✗	✗	✗	✗	✗	✗
PHE Centre 2015	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	✓	✗	✗	✗	✗
PHE Region 2016	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	✓	✗	✗	✗
CANREG 2009	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	✗	✗	✗
CALNCV 2017	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	✗	✗
CALNCV 2016	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	✗
SCN 2019	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA

Method 2: Calculating statistics using weighted data

When to apply it

This method is appropriate when the source and required areas are not coterminous or hierarchical but the counts and denominators required to calculate the statistic for the source areas are available. For more complex statistics such as DSRs, this means the age-breakdown counts and denominators required to calculate the source area statistics must be available.

Description of method

This method estimates counts and denominators for each source/required area overlapping segment and then sums these to produce estimated counts for the required areas. Statistics are then calculated in the usual way from the required area estimated counts and denominators (but without confidence intervals).

How to apply it

1. For each contributory source area group, calculate the proportion of each source area population that falls within each required area group to give weightings. Some weightings will take a value of 1 if the whole of a source area is within a single required area.
2. Multiply the counts and denominators for the source areas by the weightings calculated in step 1 to give estimated counts and denominators for each source/required area overlap combination.
3. Sum these estimated counts and denominators to give an estimated count and denominator for each required area.
4. Calculate the required statistics in the normal way using the estimated counts and denominators for the required areas.
5. As this method involves estimation, it is not appropriate to calculate confidence intervals using the standard method for the statistic.

Example with resident populations

Using this method, LTLA data from the indicator “Hospital admissions caused by unintentional and deliberate injuries in young people (aged 15-24 years)”[1] can be used to estimate the equivalent rates fir CCG resident populations. Using the proportions shown in table 1, counts for these hospital admissions, and the equivalent denominator, can be estimated for each overlapping segment between the LTLAs and CCGs (figure 5).

Weighted counts and denominators for each CCG can then be calculated by summing the weighted segment counts and denominators respectively. These weighted counts and denominators can then be used to calculate the crude rate of hospital admissions (table 5). The rate for Nottingham City CCG is the same as Nottingham City local authority as they are coterminous, whereas Nottingham North and East CCGs rate is made up of parts of three different local authorities.

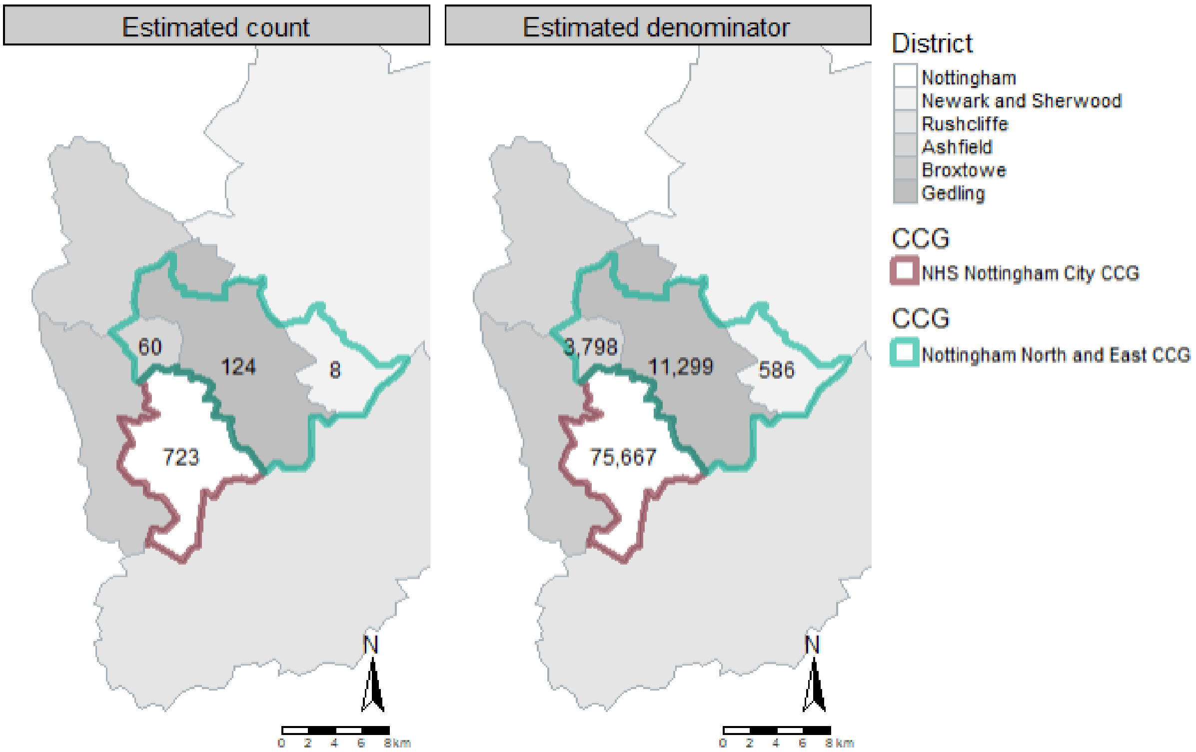

Figure 5. Estimated counts of hospital admissions caused by unintentional and deliberate injuries in young people (aged 15-24 years), 2017/18, for each segment made up from overlapping CCGs and LTLAs in Nottinghamshire.

[1]Data from PHE Fingertips indicator 90285: https://fingertips.phe.org.uk/profile/child-health-profiles/data#page/3/gid/1938133236/pat/6/par/E12000004/ati/102/are/E10000021/iid/90285/age/156/sex/4

Table 5. Estimated hospital admissions caused by unintentional and deliberate injuries in young people (aged 15-24 years) for Nottingham City CCG and Nottingham North and East CCG from LTLA data, 2017/18

LTLA	Count	Denominator	CCG	% LA population in CCG	Weighted count	Weighted denominator	Value*
Nottingham	723.0	75,667.0	Nottingham City	100%	723.0	75,667.0	95.6
Ashfield	219.0	13,754.0	Nottingham North & East	28%	60.5	3,798.9	159.2
Gedling	131.0	11,878.0	Nottingham North & East	95%	124.6	11,299.5	110.3
Newark and Sherwood	186.0	12,795.0	Nottingham North & East	5%	8.5	586.5	145.4
Estimated			Nottingham City		723.0	75,667.0	95.6
Value			Nottingham North & East		193.6	15,684.9	123.4

*Crude rate per 10,000

Example with resident and registered populations

This method can also be applied to organisational data, such as GP registered populations, and estimations between organisational and geographical boundaries. For example, using the GP by LSOA population cross-tabulation, it is possible to estimate the prevalence of depression in any given LSOA from the indicator “Depression: Recorded prevalence (aged 18+)”[2] that has data for GPs. Figure 4 (right) shows the LSOA Nottingham 030D and the surrounding GPs where the population of the LSOA are registered. The proportion of each GP’s registered patients living in this LSOA can be calculated and then multiplied by the GP’s registered count to give an estimate of the number of people registered to that GP with depression that live in the LSOA. The same can be applied to the denominator. These weighted counts and denominators can then be summed in the same way as shown in table 3 and the value calculated for the LSOA (table 6). Calculating standard confidence intervals for the aggregated LSOA value is not valid as the GP registered populations are not coterminous with the LSOA resident populations.

[2]Data from PHE Fingertips indicator 848: https://fingertips.phe.org.uk/profile/general-practice/data#page/3/gid/2000003/pat/152/par/E38000132/ati/7/are/C84039/iid/848/age/168/sex/4

Table 6. Calculating Depression: Recorded prevalence (aged 18+), 2018/19, for the LSOA Nottingham 030D resident population from GP registered populations

GP code	Count	Denominator	LSOA	% GP population in LSOA	Weighted count	Weighted denominator	Value*
C84004	590.0	5,429.0	E01013984	0%	0.0	0.2	10.9
C84023	2,566.0	42,946.0	E01013984	0%	2.4	39.9	6.0
C84039	863.0	10,157.0	E01013984	0%	3.1	36.3	8.5
C84044	425.0	7,975.0	E01013984	10%	43.6	818.4	5.3
C84065	292.0	4,355.0	E01013984	0%	0.8	11.8	6.7
C84091	622.0	5,696.0	E01013984	0%	0.2	2.2	10.9
C84092	311.0	3,459.0	E01013984	0%	0.1	0.8	9.0
C84096	75.0	2,677.0	E01013984	0%	0.2	7.7	2.8
C84107	515.0	6,591.0	E01013984	1%	5.6	72.1	7.8
C84112	178.0	2,930.0	E01013984	1%	2.5	40.5	6.1
C84117	1,505.0	19,863.0	E01013984	0%	0.5	6.0	7.6
C84122	556.0	6,571.0	E01013984	1%	6.4	76.0	8.5
C84136	284.0	3,017.0	E01013984	0%	0.1	0.8	9.4
C84647	62.0	1,243.0	E01013984	0%	0.0	0.8	5.0
C84705	566.0	4,603.0	E01013984	0%	0.9	7.1	12.3
C84714	486.0	6,411.0	E01013984	0%	0.1	0.7	7.6
Y02847	1,393.0	9,511.0	E01013984	0%	0.5	3.3	14.6
Y03124	471.0	4,089.0	E01013984	1%	3.6	30.9	11.5
Estimated value:			E01013984		70.6	1,155.5	6.1

*Prevalence

Method 3: Estimating using weighted statistic values

When to apply it

This method is appropriate when the counts and denominators required to calculate the statistic for the source areas are not available, irrespective of whether the source and required areas are coterminous/hierarchical or not.

Whilst this method is most commonly required for aggregate statistics such as DSRs, which are more likely to be published without the 5-year age band counts required for their calculation, it is also appropriate to apply this method to simpler statistics such as crude rates and proportions if the counts for these are not available.

Description of method

For values such as DSRs, more data are needed to calculate the value than just the overall count and denominator, eg counts and populations broken down by 5-year age bands. Where these breakdowns of the data are not available for the source areas, the methods described earlier cannot be used to derive estimates for the required areas. Instead, a weighting can be applied to the statistical value directly.

This uses a similar method to method 2, but instead of multiplying by the proportion of the source area population that falls within the required area, each contributory source area value is multiplied by the proportion of the required area population that is accounted for by that source area population. These values are then summed for the required area to give a weighted value. This means that source areas that make up a larger proportion of the required area will have a greater influence on the weighted value.

Where there is a one to one relationship between the source and required areas, the value will remain the same, but where the areas are non-coterminous/non-hierarchical the value will be weighted across all the source areas that overlap with the required area.

How to apply it

1. For each required area, calculate the proportion of the population that is accounted for by each contributory source area population to give weightings.
2. For each required area, multiply the statistic values for all contributory source areas by the weightings calculated in step 1 to give weighted statistical contributions for each contributory source area.
3. For each required area, sum the weighted statistical contributions to give an estimated statistical value.
4. As this method involves estimation, it is not appropriate to calculate confidence intervals using the standard method for the statistic

Example with resident populations

Using this method, local authority data for the indicator "Mortality Rate from Causes Considered Preventable"[3] can be used to estimate equivalent mortality rates for the two CCG resident populations (table 7).

Table 7. Estimated mortality rate from causes considered preventable for Nottingham City CCG and Nottingham North and East CCG from LTLA data, 2015-17

LTLA	Value*	CCG	% CCG population in LA	Weighted value*
Nottingham	263.8	Nottingham City	100%	263.8
Ashfield	204.9	Nottingham North & East	23%	47.0
Gedling	164.1	Nottingham North & East	73%	120.5
Newark and Sherwood	171.8	Nottingham North & East	4%	6.3
Estimated value:		Nottingham City		263.8
		Nottingham North & East		173.8

*Directly standardised rate per 100,000

[3]Data from PHE Fingertips indicator 92488: https://fingertips.phe.org.uk/profile/public-health-outcomes-framework/data#page/3/gid/1000044/pat/6/par/E12000004/ati/202/are/E10000021/iid/92488/age/1/sex/4

Example with resident and registered populations

This method can be applied to organisational data and estimations between organisational and geographical boundaries, such as estimating rates for GP registered populations from LSOA resident populations. As the GP populations by LSOA can be aggregated to higher registered-based organisations and higher geographical boundaries respectively, it also means that data from higher level geographical boundaries, such as UTLAs, can be used to estimate equivalent data for higher level organisations, such as STP registered populations. DSRs of hospital admissions in registered STP populations can be estimated from the equivalent rates in local authority resident populations.

For example, the proportion of Lincolnshire STPs registered population that lives in each UTLA can be calculated from the GP populations by LSOA (table 8). Most patients registered to a GP in Lincolnshire STP live in Lincolnshire local authority. There is a wide spread of a small number of patients registered to GPs in Lincolnshire STP across a number of other local authorities, which reflects the spread in GP patients' residential addresses, as shown in figure 4. The counts of these local authorities that make up very low proportions of the STP can still be used to estimate the count for the STP but will have a negligible impact on the weighted value. This third method can then be applied to the local authority DSRs to estimate the rates of mortality from causes considered preventable for Lincolnshire STP registered populations, as shown in table 8.

Table 8. Estimated mortality rate from causes considered preventable for Lincolnshire STP from UTLA data, 2015 – 17

LA	Value*	STP	% STP population in LA	Weighted value*
Lincolnshire	178.7	Lincolnshire	96.2	171.9
Peterborough	203.9	Lincolnshire	1.1	2.2
Leicestershire	151.8	Lincolnshire	1	1.4
Rutland	139.2	Lincolnshire	0.6	0.9
North Lincolnshire	190	Lincolnshire	0.3	0.6
North East Lincolnshire	232.7	Lincolnshire	0.3	0.6
Nottinghamshire	180.5	Lincolnshire	0.2	0.4
Northamptonshire	184	Lincolnshire	0.2	0.4
Cambridgeshire	148.2	Lincolnshire	0.1	0.1
Norfolk	166.6	Lincolnshire	0	0.1
Nottingham	263.8	Lincolnshire	0	0
Doncaster	216	Lincolnshire	0	0
Derbyshire	181.9	Lincolnshire	0	0
Leicester	223.6	Lincolnshire	0	0
Rotherham	208.8	Lincolnshire	0	0
Somerset	157.3	Lincolnshire	0	0
Leeds	216.7	Lincolnshire	0	0
Kensington and Chelsea	143.2	Lincolnshire	0	0
Newham	176.9	Lincolnshire	0	0
Wandsworth	167.9	Lincolnshire	0	0
Hertfordshire	147.9	Lincolnshire	0	0
Kingston upon Hull	279.8	Lincolnshire	0	0
Torbay	197.7	Lincolnshire	0	0
Birmingham	220.8	Lincolnshire	0	0
Estimated value:		Lincolnshire		178.6

*Directly standardised rate per 100,000

Caveats and metadata

Caveats

When aggregating using the appropriate methods and lookups from raw data, the aggregated values have no further caveats to those of the raw data. This also applies when aggregating from lower level areas that are coterminous with the required geography, such as aggregating local authorities together for regions.

However, when the methods outlined above are used with non-coterminous areas that use weighted proportions there are several caveats that need to be considered when using and interpreting the estimated values.

Consistency across source areas

These methods assume that the value being aggregated is consistent across the source areas used in the aggregation. When population proportions are applied to the source areas counts and denominators it splits them out by the proportion of the population also living in the required area, regardless of any variation across the source area.

For example, the method of estimating CCG counts of hospital admissions caused by unintentional and deliberate injuries in young people (aged 15-24 years) from LTLA data proportioned out the LTLA counts and denominators by the proportion of the population that lived within the CCG boundary (figure 1). This assumes that this rate of hospital admissions throughout each LTLA is consistent and the counts can therefore be split evenly by the population proportions.

However, this is unlikely to be completely true, and could be far from true. Other factors, such as deprivation, are known to affect public health measures and indicators and vary within LTLAs (figure 6). These methods therefore provide a way to apply the LTLAs rate to smaller segments of their populations for aggregation, but don’t control for any confounding variables, such as deprivation, that would influence the rate between these smaller segments of populations. If there are known reasons why particular rates will vary a great deal within individual source areas, it is probably inappropriate to use this approach in those cases.

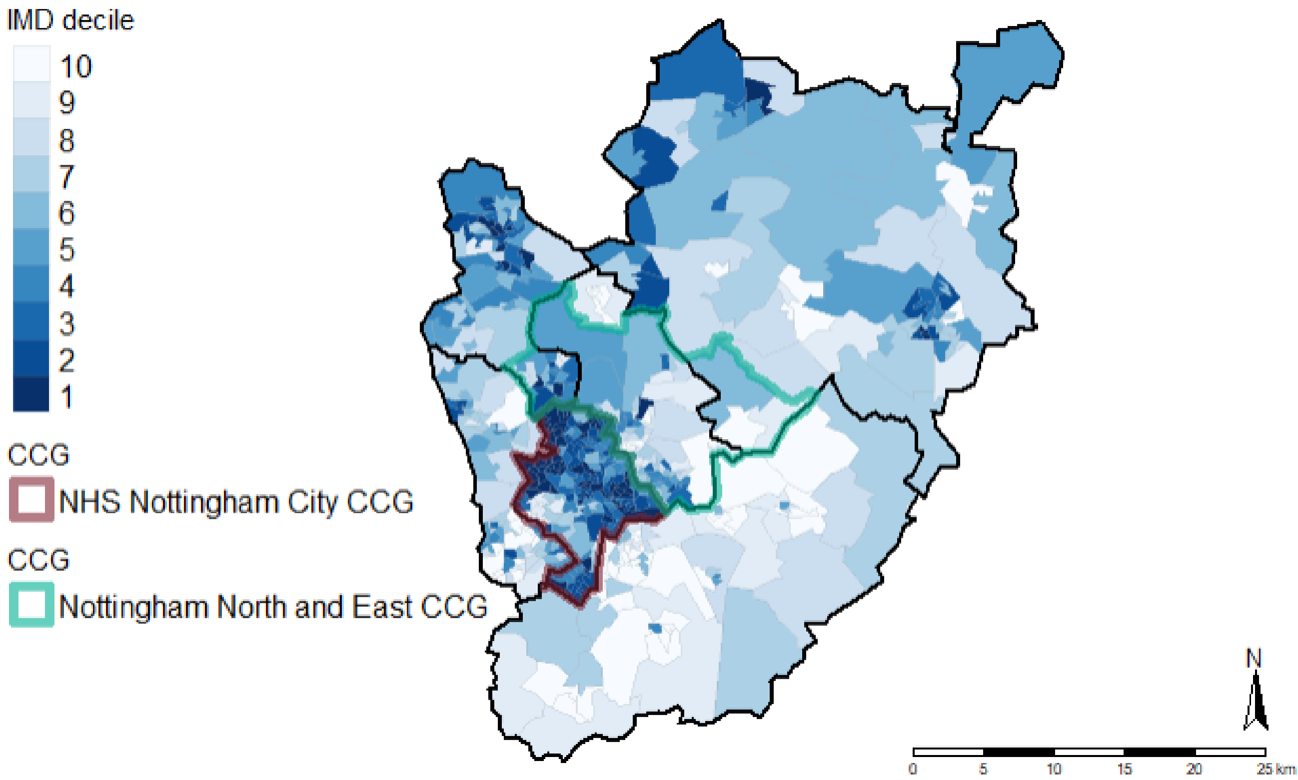

Figure 6. Map of Nottingham City CCG and Nottingham North and East CCG with overlapping lower tier local authorities and Index of Multiple Deprivation (IMD) decile.

Example metadata

Where these methods are used in analysis and published figures, this document can be referenced in the metadata to detail the method used. The following sentences can be used by adding in the relevant source area, required area and method used:

Method 1: "[required area] figures calculated by aggregating [source area] data using the first method described in the PHE aggregation guidance, weblink"

Method 2 & 3: "[required area] figures estimated by aggregating [source area] weighted data using the [method number] described in the PHE aggregation guidance, weblink"