



Ministry of Housing,
Communities &
Local Government

Title:	Needs & Redistribution Technical Working Group
Paper:	NR TWG 19-04: Children and Young People's Services (CYPS) relative needs formula
Date:	25 June 2019
Venue:	Local Government Association, 18 Smith Square, London

1. As part of the review of local authorities' relative needs and resources, MHCLG and DfE commissioned a research project to develop a new multi-level model for Children and Young People's Services (CYPS). This paper provides an update on modelling progress, based on the research project undertaken by LG Futures and academic partners.

Background

2. In order to deliver an up-to-date assessment of the relative needs between authorities providing Children and Young People's Services, we believe that the best available option is to develop a new service-specific formula which offers appropriate levels of analytical robustness. The multi-level model approach identified by MHCLG and DfE helps to account for the low incidence and high variation within children's services and will reflect the more detailed approach taken for adult social care, another high risk, complex and high-profile area.
3. The Government ran a competitive bidding process to complete this work and following initial scoping work in autumn and winter 2017/18, successful contractors LG Futures (a specialist consultancy firm) identified the preferred option to proceed with a model based on individual level child data, using data from the National Pupil Database owned by DfE. This is a highly robust approach as it looks both at the characteristics of individual children as drivers of need as well as the characteristics of the areas in which they live. LG Futures presented primary work, including the process involved and how an individual-level approach would work to the Technical Working Group in March 2018.¹

¹

https://www.local.gov.uk/sites/default/files/documents/Children%27s%20services%20research_update%20-%20March%202018.pdf

4. This uses individual child-level social care activity data, along with wider child-level data and socio-economic data about the area, to predict the likelihood that a child will have interaction with social care services. This means that the model is highly detailed, built up from individual children's information and allows the formula to go further than the traditional dependence on geographically-aggregated data.
5. Using data at this level also removes the impact of local authority (or court service) decision-making from the relative allocations. For example, if Authority A has the same number of children with characteristics that suggest they would be within the care system as Authority B, both authorities will receive the same financial allocation for this element of the model, regardless of whether these children enter the care system. This means that if a particular authority has invested heavily in targeted early intervention and has reduced need in a particular area, they will still receive the same level of funding for the children that they have in this area. Therefore, the model will not penalise those authorities that are more efficient or place more emphasis on preventative services.

Section 1: Structure of the model

6. **The first step in building the model is to combine data on the detailed characteristics of children with data on the use of children's services.** This involved linking data at an individual child level from:
 - a. the National Pupil Database (NPD) and Individual Learner Record (ILR) which details the socio-demographic characteristics of all children registered at state-maintained schools and Further Education colleges, with
 - b. the Children in Need (CIN) and Children Looked After (CLA) datasets, providing details on all contacts with children's services, (including any children not listed in the NPD/ILR)
7. This was supplemented with data collected directly from local authorities on the Lower Level Super Output Area (LSOA) of origin for all children on their CLA and CIN registers.
8. **The second stage in creating the model is to determine how best to measure service activity.**
9. The model uses a database of c.3.3m individual level child records from local authorities across the country. Combining this data with data on whether these children have accessed different service areas of children's services means that we can look at the combinations of characteristics which result in a child accessing social care. These are used to predict whether all children in the NPD are likely to have

POLICY DEVELOPMENT: NOT A STATEMENT OF GOVERNMENT POLICY

contact with social care services. These predictions are then aggregated to local authority level to produce a funding allocation model for each service area. The individual models are then weighted and a local authority level CYPS needs share is produced.

10. During construction, the model was divided into seven ‘expenditure/activity’ service areas based on data categories readily available nationally and providing a ‘best fit’ between LA aggregate spend data and social care activity data. These were:

- a. Child, young people and family support services;
- b. Safeguarding triage: assessment, case management, and commissioning;
- c. Residential Care for Children Looked After (CLA), including education of CLA;
- d. Fostering for Children Looked After, including short breaks and education of CLA;
- e. Supporting legal permanence in alternative families;
- f. Care leaver services; and
- g. Youth Justice.

11. Each of these service area categories was tested to see how they aligned with appropriate activity metrics. The activity metric is a measure of service utilisation in the dataset which will define whether an individual in the dataset has accessed one of these service areas. These activity metrics need to be available in the child-level dataset. Contractors, in discussion with officials at DfE, identified four appropriate activity metrics. These are:

- a. CIN – whether the child is registered CIN on 31st March
- b. REF – whether the child is referred with further action in the year
- c. CARE – whether the child has ever been in residential or foster care during the year; and
- d. CEASED – whether the child left care during the year for any reason.

12. Using these four activity metrics, contractors have produced model options aligning the service area categories above with the activity metrics. Our view is that a three metric model provides a good balance between accuracy and simplicity. The REF metric (whether referred with further action in the year) has a lower correlation with the service expenditure category which it acts as a proxy for than the CIN metric. We will now consider which model is most appropriate to use within the review.

Metric	Service area
CIN	1. Child, young people and family support services & 2. Safeguarding triage: assessment, case management, and commissioning & 7. Youth Justice
CARE	3. Residential Care for Children Looked After & 4. Fostering for Children Looked After
CEASED	5. Supporting legal permanence in alternative families; 6. Care leaver services

Question: Do you have views on the activity metrics used in the model?

Section 2: Characteristics driving service use

13. The individual characteristics of over 3 million children in the modelling dataset are acting as the ‘cost drivers’ in the multilevel model. Characteristics could only be used where they were present both in the modelling dataset and in the dataset of all NPD children which would be used to operationalise the model. The variables which were considered to be robust and showed a significant contribution to the model are:
- a. Sex (Male / Female)
 - b. Age (6-7 years; 8-9 years; 10-11 years; 12-13 years; 14-15 years; 16-17 years)
 - c. Ethnicity (Minor ethnic groups used, n=17 incl. unknown)
 - d. Free school meals (FSM) eligibility (yes/no)
14. Some characteristics were considered to be too variable in terms of local level application to be robust. Variables were also excluded from consideration where there were significant amounts of missing data. This was particularly the case for those looking at duration across a school year, due to the numbers of pupils moving schools mid-year.
15. In addition to the individual characteristics in the dataset, data about the Lower Super Output Area (LSOA) from which the child originates is also appended to each child. This is referred to as pseudo-individual data (attached to an individual but referring to the characteristics of the small area around the child). LSOA-level variables which were considered to be robust and showed a significant contribution to the model are:
- a. Indices of Deprivation Affecting Children (IDACI) score
 - b. Proportion of children in LSOA with limited activities
 - c. Proportion of parents in LSOA with level-1 educational qualifications or below
 - d. Proportion of overcrowded households in LSOA
 - e. Population density of LSOA

POLICY DEVELOPMENT: NOT A STATEMENT OF GOVERNMENT POLICY

16. Several variables at LA-level were also considered for inclusion alongside the child-level dataset. These were:

- a. Rates of under-18 pregnancies
- b. Rates of self-harm/suicide
- c. Rates of admissions for alcohol-related conditions
- d. Rates of juvenile convictions or cautions
- e. Net international and UK migration
- f. Rural population

17. None of these variables met the accepted criteria for improvement to the model statistics. This means inclusion of level 2 variables is likely to result in distortion of the model rather than improvement. This is not unexpected as the evidential base of over 3 million separate data points at child level has such a high level of predictive power that inclusion of a relatively small number of LA data points is unable to add significantly to this.

Section 3: Characteristics driving need

18. The model is built to describe likelihood of service usage. The advantage of a multi-level model is that it is effective in neutralising local authority service level decision-making. However, it is still important to consider whether the characteristics which drive service use also describe the need to use services. It is also worth noting that there will be some level of need that is inherently random and unpredictable, therefore we cannot expect a model with any level of complexity or number of variables to predict 100% of need.

19. Of the variables identified above, there is good evidence that most of these are key drivers for need to use services. However, there is mixed evidence with regard to ethnicity as a driver for need to use services.

20. The model currently uses White British children as a reference category, and suggests that some Black ethnic categories are more likely to be users of children's services, with Asian Indian, Asian Bangladeshi, Asian Pakistani and Chinese children less likely. There is evidence to suggest that this may represent lower need to use services for a number of reasons:

- a. More extended family structures mean need for care is absorbed within the family.
- b. There is evidence to suggest that there are lower levels of drug and alcohol abuse among these communities, although it is possible that access to these services reflects a similar pattern of lack of use of services.

POLICY DEVELOPMENT: NOT A STATEMENT OF GOVERNMENT POLICY

21. However, it is also plausible that the differences in ethnicity represent groups who are experiencing unmet need, or that entrenched patterns of behaviour on the part of children's services may result in overuse of services by some groups.
22. Given the mixed evidence in this area, there is a decision to be made as to whether we include ethnicity as a driver within the model. Although inclusion or exclusion makes relatively little difference to the model overall, it can have a significant impact for particular authorities.
23. The primary difference of excluding ethnicity from the model would be to reduce allocations in areas with high proportions of children of those ethnicities assessed as having high usage (e.g. children of some Black ethnic categories). This may be problematic as those children will already be in the system and therefore will not cease to place a cost burden on the local authority. Similarly, for those authorities who are considered to have unmet need in particular ethnic groups, it is far from certain that an increase in non-ringfenced funding will result in those authorities investing in seeking out families who do not currently have contact with the system.
24. While we accept that ethnicity provides a significant challenge in assessing need, we do not feel that ignoring this factor in allocating funding is an effective way to improve the performance of the system in the short term. In recent engagement, members of the Association of Directors of Children's Services' (ADCS) Resources and Sustainability Policy Committee found it logical to include ethnicity in the new formula given that it does drive current usage.
25. **Balancing the known impact of ethnicity on usage of services against the possible impact of ethnicity on the supply of services, we are minded to include ethnicity in the model, subject to further analysis, including the implications for equalities.**

Question: Do you have views on inclusion of ethnicity within the model?

Section 4: Updating the model

26. As the model does not use population data, we cannot use population projections in order to forecast relative needs in the way we are for most other models within the review.
27. It may be the case that the most appropriate option is to fix the formula until an appropriate point to update the model is reached e.g. when carrying out a full refresh. It is possible to update the model without rebuilding, as has been done with the current formula.

Question: Do you have views on updateability of the model?

Section 5: Next stages

28. As the model progresses to completion, we will consider consultation and engagement with the sector. **We would welcome your views on how best to achieve this.**