Technical Annex on the Children and Young People Services (CYPS) Relative Needs Formula (RNF)

Introduction

The Children and Young People's Services (CYPS) formula will help to distribute funding to local authorities for the provision of children and young people's services. This includes spending on children's social care such as the cost of children living in residential and foster care.

The formula works by assessing the relative need of a local authority, compared to others, by estimating the likelihood that each child in a local authority will use children's social care, by either being a child in need, in care, or having ceased being in care. It does this by examining which child and neighbourhood characteristics are most associated with a child interacting with the children's social care system.

Methodology

Constructing the data:

The first step in building the model was to combine (1) data on the detailed characteristics of children with (2) data on the use of children's services. This involved linking data at an individual child level from:

- **2022-23 and 2023-24 National Pupil Database (NPD):** This database provides detailed information on the socio-demographic characteristics of children attending:
- **Schools:** As detailed in the School Census
- **Colleges:** As detailed in the Individualised Learner Records (ILR)
- **Alternative providers:** As detailed in the Alternative Provision Census (APC)
- **2023-24 Children in Need (CIN) Census:** This dataset includes information on vulnerable children referred to social care services because their health or development is at risk. It covers those receiving support from social care services and children subject to child protection plans.
- **2023-24 Children Looked After (CLA) Census:** This dataset provides information on children under local authority care. It records data on every child looked after by a local authority at any time during the year.

By combining the CIN and CLA datasets with the National Pupil Database, we obtained comprehensive details of all contacts with children's services, including any children not listed in the National Pupil Database.

This combined dataset is supplemented with data showing the Lower layer Super Output Area (LSOA) of origin for all children on the CLA and CIN registers. This allows

the model to consider the characteristics of the areas individual children are from, rather than where they are placed.

The use of individual and LSOA level data means that the model is more detailed and better at capturing variation compared to a model built using local authority level data. Using data at individual level also removes the impact of local authority decision-making (or that of other agencies, such as the court service) from the relative shares the model produces.

Building the model

Using multi-level modelling techniques, we identified the combinations of characteristics that best predict whether a child in the dataset is likely to access social care in one of three ways, the dependent variables:

- Being registered as a Child in Need on 31st March of a given year.
- Being a Child Looked After (in residential or foster care) during the year.
- Having ceased care during the year for any reason during the year.

The explanatory variables considered to be robust and included in the model are:

- Sex of child (categorised as male or female)
- Age of child
- Eligibility for free school meals (FSM) on date of the census

In addition to the individual characteristics in the dataset, data about the Lower layer Super Output Area (LSOA) from which the child originates was also then appended to each child. The LSOA-level variables which are considered to be robust and showed a significant contribution to the model are:

- Socio-economic deprivation level in child's LSOA (as measured by the 2025 Income Deprivation Affecting Children Index)
- Proportion of children in child's LSOA with poor health
- Proportion of overcrowded households with dependent children in child's LSOA
- Population density (measured in persons per km2) in child's LSOA
- Travel time from LSOA centroid to nearest town centre (mins)

Two first-order interaction effects were also included in the final model:

- Interaction between FSM eligibility and child being aged 16 or 17
- Interaction between FSM eligibility and socio-economic deprivation level in child's LSOA (as measured by the IDACI score)

Several variables at local authority level were also considered for inclusion alongside the child-level dataset. However, none met the accepted criteria for improving the model fit, as they were more likely to distort the model rather than enhance it. We note that this is unsurprising given the predictive power of an evidential base comprising over 7 million child-level data points is so great, that there is no significant benefit from including a relatively small number of around 150 local authority-level data points.

Functional form of the model

Regression modelling is used to derive relationships for the prediction of a dependent variable in terms of other variables. In this case, the three dependent variables, which are 1) registered as a Child in Need, 2) Being a Child Looked After and 3) Having ceased care during the year are all binary. A logistic regression is needed to fit the relationship between a binary outcome and a set of explanatory variables into a linear regression framework.

The overarching CYPS model is made up of three separate multilevel logistic regression models. There is a separate logistic regression for each of the dependent variables. The explanatory variables in each logistic regression model are the same and are the variables listed in the previous section.

This is the general form of a logistic regression:

$$\text{logit}\, p_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots \beta_r x_{rij} + \beta_{r+1} x_{r+1j} + \beta_{r+2} x_{r+2j} + \cdots \beta_{r+s} x_{r+sj}$$

where:

p_{ij} = the probability that child i originating in LA j will report a positive outcome with respect to the criterion of interest;

logit refers to the logistic transformation: $ln \frac{p}{1-p}$;

β_{0j} is a randomly varying-intercept, with $\beta_{0j} = \beta_0 + u_{0j}$; with β_0 a constant to be determined and u_{0j} the variance at LA-level; with $u_{0j} \sim N(0, \sigma_{0r}^2)$;

$x_{1ij}, x_{2ij}, \ldots x_{rij}$ are *r* Level-1 variables included in the model;

$x_{r+1ij}, x_{r+2ij}, \ldots x_{r+sij}$ are *s* Level-2 variables included in the model;

$\beta_1, \beta_2 \ldots \beta_{r+s}$ are parameter estimates of the Level-1 and Level-2 variables to be determined by the modelling process.

The parameter estimates used in the model are listed in Table 1. Positive numbers mean that this variable increases the probability that a child will interact with social services, whereas negative numbers mean this variable decreases that probability.

Table 1: Parameter Estimates for the Children and Young People's Services Formula

Model parameter estimates			
	Child in Need	Child Looked After	Ceased Care
Intercept	-5.2823	-7.2105	-8.2758
Free School Meals eligible (FSM)	2.2679	2.4190	2.5209
Gender (ref='Male')	-0.1171	-0.0917	0.0209
Age 8-9	0.0117	0.0952	-0.2331
Age 10-11	0.0799	0.2215	-0.4116
Age 12-13	0.2527	0.4438	-0.2873
Age 14-15	0.4323	0.7264	-0.0117
Age 16-17	0.5978	1.0824	0.4709
Income Deprivation Affecting Children Index (IDACI)	2.5444	3.8912	2.4879
Poor health	5.1238	5.0618	0.6947
Overcrowded households	-0.6764	-0.3659	-0.4297
Population Density	0.0000	0.0000	0.0000
Travel time	-0.0065	-0.0135	-0.0198
FSM*Age 16-17	-0.3075	-0.3802	-0.4663
FSM*IDACI	-1.7723	-2.9115	-1.8687

Model outputs:

The model predicts the likelihood of whether each child in the country will be:

- Registered as a Child in Need (CIN) on 31st March of a given year.
- A Child Looked After (CLA) (in residential or foster care) during the year.
- Ceased care during the year for any reason during the year. (Ceased)

These estimates are then aggregated to the local authority level. The numbers of CIN, CLA and Ceased are then converted to a local authority share of the total estimated number of CIN, CLA and Ceased.

Applying a weighting

We then apply a resource weighting to each local authority's shares of CIN, CLA and Ceased to ensure that the final relative need share estimates for each local authority reflects the relative resource needs of the different areas of CYPS activity.

The different activity metrics (that is CIN, CLA and Ceased) are weighted using the three year average of national expenditure in the corresponding section 251 expenditure lines. The mapping between activity metrics and spending lines were determined by those which had the best correlation with the available activity metrics. The shares which each authority would receive for each of the three separate activity metrics are multiplied by the weights (as in the table below) to create a weighted share.

Table 2: Activity metrics which proxy each of the Section 251 service areas

Metric	Service area	Resource weight
CIN	(1) Child, young people and family support services, (2) Safeguarding triage: assessment, case management, and commissioning, (7) Youth justice	48.7%
CLA	(3) Residential Care for Children Looked After & (4) Fostering for Children Looked After	40.6%
Ceased	(5) Supporting legal permanence in alternative families; (6) Care leaver services	10.6%

The original report written by academics that designed the formula Children and Young People's Services Formula Review.

Data sources for the formula's variables

Table 3: Data sources for each of the formula's variables

Formula variables	Data source
Free School Meals	National Pupil Database
Gender	National Pupil Database
Age	National Pupil Database
Income Deprivation Affecting Children	Income Deprivation Affecting Children Index, part of the English Indices of Deprivation 2025
Poor health	Census 2021

Overcrowded households	Census 2021
Population Density	Census 2021
Travel time	Journey time to key services by lower super outputs area 2018

Changes from the Fair Funding Review consultation

Variable Changes Implemented

- **Updating the Income Deprivation Affecting Children Index (IDACI)**;

Update with the 2025 release, in keeping with using the most up to date data available. This updated index also addresses the consultation feedback comprehensively. The 2025 IDACI is now designed to reflect housing costs and incorporates data on all benefits claimants, critically including those claiming Universal Credit. This shift ensures the formula captures the modern landscape of benefit receipt and provides a more comprehensive measure of income deprivation experienced by children.

- **Refinement of the overcrowding variable**;

The original variable specification included a general measure of overcrowded households. Following further testing and analysis, the metric was refined to improve its analytical relevance to children's social care services. The change involved replacing the previous "Proportion of overcrowded households" with the more specific measure: "Proportion of households with dependent children which are overcrowded". This targeted variable optimises the variable's contribution to the predictive accuracy of the formula, focusing the housing stress indicator precisely on the population segment relevant to the formula's outcome variable.

- **Removal of the parental qualifications variable**

The parental qualifications variable, initially included in earlier iterations of the model, was permanently removed from the formula. This decision was made in accordance with best practices in modelling and simplification. Analytical testing confirmed that the removal of this variable actually improved the model's overall ability to predict need. This suggests that the explanatory variance previously attributed to parental qualifications was being more effectively and fundamentally captured by other, more robust deprivation and contextual measures already included in the multi-level architecture, such as the refined 2025 IDACI. Removing the variable decreased model complexity while simultaneously enhancing its statistical efficacy.