Peer Review of: The Children and Young People's Services Formula Review

February 2025

Author: Professor Anna Vignoles

Contents

Overview

LG Futures, the University of Huddersfield and the University of Plymouth were commissioned by the Ministry of Housing, Communities and Local Government (MHCLG) and the Department for Education (DfE) to develop an updated and rigorous model to underpin the allocation of resources to Local Authorities for Children and Young People's Services (CYPS). The work is part of a broader set of projects informing the Review of Relative Needs and Resources (RRNR).

LG Futures and their collaborators were asked to take account of the basic principles for designing a new relative needs assessment methodology namely:

- Simplicity

- Transparency

- Contemporary

- Sustainability

- Robustness

- Stability

LG Futures proposes a model that could replace the underlying assessment of relative needs and resources currently being used as the basis for funding Local Authority baselines, which is more than a decade old.

Here I provide peer review of the proposed methodological approach and comment on the specific assumptions that lie behind the model. Where possible I provide commentary on the appropriateness of the data used in the model, particularly on the issues of quality and fitness for purpose. However, I am not able to get sight of all of the many data sets that underpin the model and hence cannot provide quality assurance on each of them. This scrutiny report also does not address any issues relating to the share of spending that should be allocated to CYPS.

In summary, the LG Futures report is extremely thorough. It carefully documents the existing funding system and spells out exactly how the proposed funding methodology has been developed. It is very clear about the assumptions on which the proposed model is based and does an excellent job of detailing the data needs (and limitations) of the model.

In terms of the principles discussed above, it appears to fulfil at least 5. The exception may be the first principle: the model is far from simple. This is acknowledged by the authors and they make a clear case as to why the complexity of the model is a price worth paying to get greater robustness. The way that the model has been constructed should however, enable more regular updating and hence it will be more transparent, contemporary and sustainable than the current system. In respect of the contemporary

nature of the data that feed the model, it will be more obvious when data has become out of date (e.g. it is clear where the model relies on 2011 Census data). In terms of stability, the model would be sensitive to dramatic changes in the underlying demographic characteristics of populations between areas. This is as it should be, in the sense that if a LA is experiencing a particularly major shift in its underlying population demographics then its funding needs to quickly reflect that. There is one potential issue with regards to stability in respect of the use of the Free School Meals indicator, which is discussed below.

It is worth noting that, irrespective of the complexity of the model, it will be very sensitive to the resource weights that are attached to each CSC activity metric used in the formula. These resource weights are potentially, as recognised in the report, a policy decision by MHCLG/DfE and it is important in terms of transparency that this is very clear in this report, and all future documentation.

It is also worth noting that the new model may produce a significant shift in the allocation of resources across LAs, at least it may be major shift for some LAs (I do not have the underlying data so cannot see the extent to which allocations may change dramatically). Given that the model is now driven by a different selection of variables, it is to be expected that significant changes in allocations may occur. There is no commentary in the report I believe on managing such a transition. This issue needs to be addressed if the formula is adopted.

Overall, I am satisfied that what is proposed meets the remit set by MHCLG and DfE for a genuinely needs based prediction of what the likely CYPS activity is for each LA and hence provides a basis for allocating the total sum allocated to CYPS across different Local Authorities. The approach appears sound. However, there are some issues which merit further consideration.

Approach

The current methodology used to allocate funding to Local Authorities consists of distinct formulae for three sub allocations, namely: Children's Social Care, Youth and Community, and LA Central Education Functions. The latter function was excluded from the remit of the LG Futures report after discussion with MHCLG and DfE. It is also important to note that Early Intervention Funding was incorporated into each LA's baseline in 2013/14 and hence is within scope of the report. The LG Futures report therefore proposes a methodology for calculating LA funding for Children's Social Care and Youth and Community functions only, including Early Intervention Funding.

Whilst elements of the formulae behind the current allocations to LAs have been updated over time, most relate to the period 2013/14 or earlier. Hence it is clear that an update is indeed necessary.

The current funding formula is based on a multi-level model of the variation in per capita costs of CSC services using postcode district level data. The underlying cost and service use data are from a 2003 Children in Need survey and the model is driven by 4 factors:

- % of children in fairly good/bad health

- % working age adults receiving IS or IS based Jobseeker's Allowance

- % children in out of work families receiving tax credit

- % children in black ethnic groups.

The LG Futures report suggests a different approach based on individual level data on need. Since needs arise at the level of the child or young person, rather than at the area level, this is a more robust approach. Aggregating predictions of individuals' needs will lead to area estimates of need that better reflect differences in populations, as compared to an approach that is based on estimating needs from the average characteristics of areas.

LG Futures proposes a single formula to determine funding allocations for CYPS at the Local Authority level, based on the modelled / predicted need of each LA. The formula would be derived from a 2 level multi-level model that uses individual level data to predict the child's social care needs (CSC), and specifically whether the child is likely to be classified as Children in Need (CiN) or Children Looked After (CLA). A 3 level model (individual, LSOA and LA) was considered and dismissed, largely on grounds of computational complexity. Estimates of need are then translated into likely level of activity at the LA level, which can in turn drive the allocation of resources.

The proposed approach requires good quality individual level data on numbers of children by age and their demographic characteristics that predict their need for CYPS. The decision to use the National Pupil Database as the source of much of the data is a sound one, given the relative robustness of that data set. Issues relating to the problems that arise from relying on this data set are discussed below.

A key principle behind this approach is that LAs should be funded on the basis of the underlying needs of their populations, i.e. as predicted by demographic risk factors. They should not be funded on the basis of how efficiently they currently spend their resources and hence their current levels of spend or activity. Hence factors that relate to differences in LA policy are deliberately excluded from the model, even though they might well predict whether a child receives a service. For example, a LA with an excellent early intervention policy may have fewer Children Looked After. LAs should not be punished for this efficient early intervention policy by being funded at a lower rate. This is a very sound principle.

Issues for consideration

The report helpfully sets out some principles and assumptions that lie behind the methodology and this highlights the main issues for consideration that come out of the report. I only focus on issues that may be problematic or require further thought.

There is inadequate data on the number and CSC needs of the 0-5 year old population since the National Pupil Database (NPD) coverage is poor for this group.

- The needs of the age 6-10 year olds in NPD are used to simulate the needs of the age 0-5 year olds. The report presents data on the correlation of service use across age groups to justify this.

- On the basis of a high correlation between age 0-5 service use and age 6-10 service use, this appears to a reasonable assumption in the absence of adequate data on 0-5 year olds. However, it is far from ideal and developing a data source for the future might be something to consider, perhaps from medical records.

A significant proportion of children are missing from one of the key data sets on which the model is built (NPD). This is largely due to children being educated in the independent school sector.

- Around 7% of children age 5-17 are missing from NPD, consistent with estimates of the size of the independent school sector.

- The report presents evidence that the vast majority of the children missing from NPD are likely to be in the independent school sector and hence it is assumed that they will have low levels of need.

- The missing children are not ignored. Instead the characteristics of the missing children are simulated by assuming that they are the same as the children for whom data is observed in a specific LSOA.

- Children being home schooled are potentially missing. This phenomenon has grown significantly in recent years. Home schooling might be correlated with CSC need.

- The % of children missing from NPD does vary hugely by LA. So this problem will affect some LAs more than others, as is documented in the report.

  - **RECOMMENDATION 1**: Investigate home schooling data held by DfE to determine the significance of this phenomenon for the model. Determine whether the Millennium Cohort Study has data on those who are not in the NPD which could be used to check the assumptions behind the simulations used. An alternative approach would be to change the parameters to test the sensitivity of allocations. For example, what material difference would it make if one assumed that a larger proportion of the missing children had higher levels of need.

The National Pupil Database will generally hold data on where the child is living, rather than their location prior to being looked after. Separate data is needed to determine where such children were originally living and hence in which LA their needs arose.

- This data was collected from 71 LAs.

- The report provides descriptive statistics which suggest the LAs that provided data were representative of all LAs.

- There is a risk that LAs that did not provide data are unobservably different from those that did. However, it is likely that any unobserved drivers of whether a LA responded are related to LA operations/efficiency/policy rather than differences in need, so on balance this data seems adequate.

- It is not clear how, going forward, one would update this information to keep the model current.

  - **RECOMMENDATION 2**: if this new model is adopted, the Departments might want to determine whether to collect this core data on a routine basis.

Cost data by type of CSC activity is complex and sometimes aggregated across different activities. Calculating both the activity rates for CSC services and the related expenditure categories required a lot of assumptions.

- The current source of data is very old (2003 Children in Need Survey) and could not be used.

- Potential source data are carefully considered in the report, including a proposal to obtain individual child level data on CSC use from each LA. These alternative options were dismissed either on the basis of coverage or practicability.

- The authors therefore rely on actual use of CSC services as recorded in the CiN and CLA data sets. This is of course a subset of total activity and it is possible that for some LAs this approach may under or overestimate their activity /need. However, in the absence of better data, this approach seems reasonable.

- Data on use of services is then mapped (with some difficulty) to s251 returns on the proportion of expenditure associated with different CSC activity categories. Again, there are issues about the data here, particularly how one allocates expenditure to a particular activity. The report appears to have done a thorough job in exhausting the possible ways in which this data might be improved.

- Overall, it is not clear how the data source used for this part of the process could be improved without considerable resource being spent on further data collection.

There are some additional issues that arise in relation to the choice of predictors in the model.

- The report carefully documents how individual, LSOA and LA predictors of need are selected, and the approach appears broadly sound. However, the final model

is sparse with few covariates (at individual level it is just age, sex, ethnic group and FSM eligibility). A number of variables are not included basically on the grounds of the poor quality or complexity of the data, as well as lack of statistical significance in the modelling. This means some variables one might expect to see are not in the model. Normally one might prefer a theoretically driven selection of covariates, but the authors note the difficulty due to the size of the data and complexity of the model and hence have gone for a parsimonious approach. This is reasonable, though as better data becomes available one might want to do robustness checks of the sensitivity of results to inclusion of key variables that are justified theoretically.

- The covariates that are included in the model differ quite significantly in substantive nature from those used in the current model.

- At LSOA and LA level, data on the "structural drivers" of need are limited to IDACI and the proportion of children in the LSOA who have parents with low qualifications. No LA variables are included in the model at all.

- One variable clearly missing from the model is the proportion of individuals who are on benefit, though of course the IDACI composite measure does include information about individuals on different types of benefit. I take the point about Universal Credit not being rolled out yet and hence this is difficult to capture. However, omission of data on UC benefit receipt at LSOA level may or may not be problematic in the future (the current model does include this). There is also no data on the proportion of individuals with higher levels of education. There is a comment that POLAR is no longer being used by HEFCE. However, POLAR has been updated and is being used by HESA I think, which might merit further investigation.

    - **RECOMMENDATION 3**: Investigate possible use of POLAR. Update model when full coverage of UC is achieved. Provide some commentary on how the covariates differ from those in the current formula model.

- A big driver in the proposed model is current FSM eligibility. The current FSM indicator variable will change over time (as families move into and out of benefits) and over the economic cycle. This means that the apparent needs of a LA could vary over the economic cycle even if the underlying population does not change. This might be correct if we think that having a material change in circumstances that causes a child to become FSM eligible also causes them to have increased need. If, however, we think long term poverty is a stronger driver of need, as compared to temporary reductions in income, then the model might be more sensitive to changes in FSM status than we would like.

    - **RECOMMENDATION 4**: Investigate whether "ever FSM" is a feasible alternative to use in the future.

- Inclusion of ethnicity as a driver of need is discussed. It may be that lower service use by some ethnic minority groups reflects bias or lack of access to services.

This would suggest it should not be included. However, one could equally make the argument in reverse for FSM eligibility. Perhaps high use of services by FSM eligible children may not reflect their needs so much as "overuse" of services because negative judgements are made about the risks faced by children living in poor households. These are problematic issues that cannot be easily resolved. I agree that the decision to retain ethnic group is on balance acceptable.

Technical issues

The model is a standard multi-level model with random intercepts. This assumes that the variation across LAs is in the intercept not the slopes and hence that relationship between the predictors and the outcomes is the same across LAs. This appears to be an assumption rather than tested.

- **RECOMMENDATION 5**: Further discussion is needed of any robustness checks on this point.

Para 9.4 is incorrect, I think. It does not imply a causal relationship between LSOA and outcomes. It could be that there are omitted individual level variables that are correlated with LSOA and hence LSOA is simply a proxy for this.

Para 9.67 refers to informal Bonferroni corrections and I was unclear what was meant by this.

What is the interpretation of the very small magnitude of the FSM/SES interaction as an odds ratio?

Paras 10.10-10.14 set out the issue of the "Unknown" ethnicity. Rather than assume that "Unknown" children are White British, could one not randomly assign an ethnicity to reproduce the proportion of the different ethnicity groupings in the LA?

- **RECOMMENDATION 6**: These technical issues would benefit from further clarification and consideration in the final report.

Conclusion

Overall I am satisfied that what is proposed meets the remit set by MHCLG and DfE for a genuinely needs based prediction of what the likely CYPS activity is for each LA and hence provides a basis for allocating the total sum allocated to CYPS across different Local Authorities.

In terms of the principles of the review, it appears to fulfil at least 5. The exception may be the first principle: the model is far from simple. This is acknowledged by the authors and they make a clear case as to why the complexity of the model is a price worth paying to get greater robustness.

In conclusion, the underpinning methodology that has been constructed to drive the model appears sound. It is worth noting however, that the authors stress the point that regular updates of data will be needed to keep the model current. This is required not just for the estimates of the denominator (population shares) but also for the prediction of need. This suggestion is vitally important, and a schedule of updates might be beneficial.

![Department for Education logo]

For any enquiries regarding this publication, contact www.gov.uk/contact-dfe.

This document is available for download at www.gov.uk/government/publications.