

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/131921/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Wilkins, David ORCID: <https://orcid.org/0000-0003-2780-0385> and Antonopoulou, Vivi 2020. Do performance indicators predict Ofsted ratings? An exploratory study of children's services in England. *Journal of Children's Services* 15 (2) , pp. 45-59. 10.1108/JCS-07-2019-0035 file

Publishers page: <http://dx.doi.org/10.1108/JCS-07-2019-0035>
<<http://dx.doi.org/10.1108/JCS-07-2019-0035>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Do performance indicators predict Ofsted ratings? An exploratory study of children's services in England

Abstract

Children's Services in England are inspected by Ofsted (the Office for Standards in Education, Children's Services and Skills). Each local authority is given a singular overall rating – outstanding, good, requirements improvement or inadequate. These ratings carry immense significance. Persistently inadequate authorities are liable to have legal responsibility for providing services outsourced to another organisation. It used to be said that Ofsted focused too much on procedural compliance and previous research has found that deprivation and spending were significantly associated with Ofsted ratings. More recently, Ofsted has introduced a new inspection framework, through which they aim to put the experiences of children at the heart of their approach. We report an analysis of forty-five variables in relation to children in need, children in care and young adults with care experience. We considered whether Ofsted inspection results were associated with reliably good or poor performance in relation to these variables. In our statistical analyses, we found no consistent patterns of difference between authorities in relation to their Ofsted rating and deprivation was the best predictor.

Keywords: social work; children and families; inspection; Ofsted

Introduction

It is widely believed by politicians and policymakers in England that Ofsted ratings of children's services are important and meaningful. When authorities are given positive ratings, this is viewed as evidence that the services provided are effective for children and families. When authorities are given poor ratings, this is viewed as evidence that services are ineffective. According to former Conservative Prime Minister David Cameron, "*We will not stand by while children are let down by inadequate social services*" (HM Government, 2015, unpaginated). In recent years, several authorities in England have had legal responsibility for providing these services outsourced as a result of persistently poor Ofsted ratings.

It used to be asserted that Ofsted's inspections were overly focused on procedural compliance (*The Guardian*, 2015). More recently, the then-head of Social Care for Ofsted, Eleanor Schooling (2017a), argued that a new inspection framework would have at its heart "*children and their experiences*". Schooling (2017b) also argued that "*process still matters in*

social work – but not as much as outcomes”. These quotes suggest a shift in Ofsted’s thinking, and an acknowledgment that previous inspection frameworks were too focused on process.

All of which hints at some very positive developments, given the negative publicity that Ofsted has often received (Shoesmith, 2016, p. 112 and p. 197; Rogers, 2017; Matthews, 2018). And yet there is a dearth of research about Ofsted and children’s social care, particularly when compared to the wealth of research about Ofsted and schools. Given the political and practical significance of Ofsted’s ratings (Jones, 2015), this is far from ideal. In this paper, we present the results of a new analysis of Ofsted ratings and their association with a series of forty-five key variables. We hope by so doing to make a modest contribution to the ongoing debate about how best to define good practice and outcomes in child and family social work (Forrester, 2017).

Brief literature review

Ofsted are responsible for inspecting a range of establishments within the education and social care sectors in England. The public (in relation to schools) and politicians (in relation to schools and social care) place a great deal of trust in their findings. Ofsted’s ratings are the second most important factor for parents when choosing a new school, behind proximity (Richmond, 2019). For children’s services, poor Ofsted ratings can destabilise the workforce and cause or contribute to a cycle of decline (Impower, 2015). On the other hand, when an authority is rated positively by Ofsted, they are likely to receive positive political and media recognition (Stevenson, 2016).

In relation to school inspections, Richmond argued that Ofsted “*has not published any research to support the notion that their judgements in schools accurately reflect the quality of education a school provides*” (p. 1, 2019). The same can be said in relation to children’s services. The National Audit Office (2018) has also found that “*Ofsted [is] unable to demonstrate that its inspections of schools represent value for money*”. Various other studies have found a range of similar problems, including the negative - or at best neutral - impact of Ofsted inspections on exam results (Shaw et al, 2003; Rosenthal, 2004) and the unreliability of Ofsted’s ratings (Campbell and Husbands, 2000).

For children’s services, we have less evidence to draw upon. Hood et al (2016) found that only a small number of performance indicators were sufficient to predict Ofsted ratings. La Valle et al (2016) also considered the relationship between Ofsted ratings and selected performance measures and found that two of the poorest authorities (as rated by Ofsted)

performed relatively well, while one of the best authorities (as rated by Ofsted) performed very poorly. Munro (Ofsted, 2014) also raised concerns about the demands placed upon authorities by the inspection process and queried whether Ofsted's judgements were reliable or valid.

In their 2016/17 annual report, Ofsted seemed to acknowledge a potential link between deprivation and their inspection ratings and committed to undertaking their own research (p. 70). However, a decision was subsequently taken not to proceed (Ofsted, 2019, personal communication).

Many of these reports and studies relate to Ofsted's previous inspection framework, which has now been replaced with a revised version known as ILACS (Inspecting Local Authority Children's Services). This new framework says there should be "*less focus on process and a greater focus on social work practice*" and that inspectors should spend "*80% of their time talking to social workers and directly observing practice*" (Ofsted, 2019a, p. 3 – 4). Ofsted also state their intention to conduct interviews with children and parents, as a way of helping triangulate their findings (ibid, p. 69). By so doing, Ofsted aims to understand not simply what local authorities are doing (in terms of process and procedural compliance) but the impact they are having on the lives of children and young people. These are welcome changes to the framework – at least according to Ofsted's own evaluation (ibid, p. 5). In an independent evaluation, Ferguson, Gibson and Plumridge (Ofsted, 2019a) draw different conclusions. For example, Ferguson et al found that "*the voices of children in need of help and protection are completely absent*" and that "*little, or no, effort is made to take into account the experiences of parents*" (p. 46). They also found that although "*what social workers said to inspectors certainly mattered...inspectors still placed huge emphasis on what could be ascertained from the administration and documentation of social work practice. Inspectors kept coming back to whether things were written down or not*" (p. 50). From this brief literature review, we identified the following three research questions:

1. How do Ofsted ratings of Children's Services relate to outcomes for children in need, children in care and young adults with care experience?
2. How do Ofsted ratings of Children's Services relate to procedural compliance?
3. How do Ofsted ratings of Children's Services relate to patterns of spending, deprivation and the characteristics of local populations?

Method

To address these questions, we obtained data about all local authorities inspected under the ILACS framework between January 2018 and March 2019 (n=47). For each authority, we obtained a copy of their most recent inspection report and identified the overall rating, and the sub-category judgements for children in need of protection, looked after children and leadership (Figure 1). We placed each authority into one of two categories based on their overall rating (Table 1).

Figure 1: Overall Ofsted ratings (%) for 47 authorities inspected in 2018-19 under the new ILACS framework

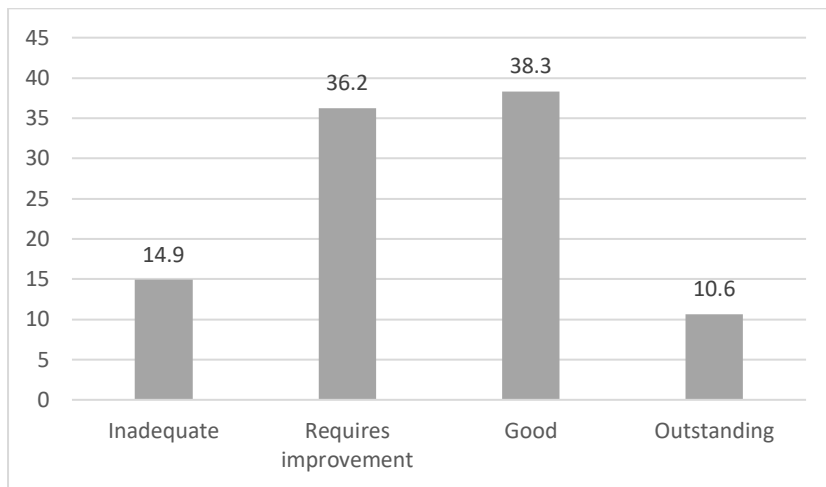


Table 1: Local authorities grouped into categories of either low or high Ofsted ratings

	Frequency	Percent
Low Ofsted ratings (inadequate and requires improvement)	24	51.1
High Ofsted ratings (good and outstanding)	23	48.9
Total	47	100.00

We grouped together authorities rated outstanding and good and those rated inadequate and requires improvement in order to create two roughly equal size categories. Having balanced categories aids with the statistical analysis that follows. We could have kept the ratings separate, but this would have increased the risk of false positive and negative findings due to the small sample size of the outlier groups (n=7 for inadequate and n=5 for outstanding). Our use of these combined categories should not be mistaken as implying that we view the categories as somehow equivalent.

Measures of spending and deprivation

Estimating spend on Children’s Services is not straightforward (Bywaters et al, 2015 and 2017). For this study, we used the same method as in our previous paper (Authors Own), providing the advantage of consistency if nothing else. We obtained section 251 returns for each authority, providing budget figures in relation to under-5s spending, safeguarding and family support (children in need and child protection), children in care and ‘other’. We divided these figures according to the local child population, giving estimated ‘spend per child’ figures for total spending (in relation to the overall child population), under 5s spending (in relation to the overall child population), safeguarding (in relation to children in need) and children in care (in relation to children in care). We acknowledge the limitations of this approach, not least because authorities may report similar types of spending within different categories. Authorities will also target their spending on different groups of children (and not the total population per se). Nevertheless, these estimated figures do provide a way of comparing between authorities of varying sizes. This approach is broadly similar to that used by other researchers in the field (Bywaters and Webb, 2018). For deprivation, we used the most recently available figures from the English Indices of Deprivation at local authority level. We ranked the authorities from most to least deprived, and sorted each authority into one of three categories – high, moderate or low deprivation (Table 2).

Table 2: Local authorities grouped into high, moderate or low deprivation.

	Frequency	Percent
High deprivation	16	34.0
Moderate deprivation	15	31.9
Low deprivation	16	34.0
Total	47	100.00

Key variables in relation to procedure and outcomes

Via the Department for Education, we obtained publicly available data on a range of variables in relation to children in need, children in care and young adults with care experience (Table 3). The most recent dataset available at the time of writing covers 2017 / 18. We selected variables in relation to procedural performance or outcomes whenever it seemed possible to compare meaningfully between different authorities (i.e. where the data

were proportional rather than raw numbers and where there were not large data gaps). This resulted in a list of variables that were both comparable between authorities and complete (or near-complete) across our sample, many of which have also been considered significant within the wider literature (Bilson and Martin, 2017). We arranged these variables into the following categories – features of the local population (e.g. the number of children in need per 10,000), children in need procedures (e.g. the timeliness of assessments), children in need outcomes (e.g. educational performance), children in care procedures (e.g. up-to-date health assessments), children in care outcomes (e.g. the stability of placements) and leaving care outcomes (e.g. the suitability of accommodation). We found no variables in relation to leaving care procedures. We also grouped the variables together as being broadly ‘procedural’ or broadly ‘outputs and outcomes’. There is no agreed standard for what qualifies as an ‘outcome’ compared to an ‘output’ in children’s services. We have used our best judgement while accepting that others may define these categories differently.

Table 3: A list of the variables used for this study, grouped into categories.

Population variables		<ol style="list-style-type: none"> 1. Referral rate per 10,000 2. Child in need rate per 10,000 3. Child protection rate per 10,000 4. Children in care per 10,000 5. Local authority deprivation 6. Spending data
Children in need (CIN) variables	Procedural	<ol style="list-style-type: none"> 7. Assessments overdue (%) 8. Child protection reviews overdue (%)
	Outputs and outcomes	<ol style="list-style-type: none"> 9. Repeat referrals (%) 10. Children in need ‘persistently absent’ from education (%) 11. Children in need with a fixed term exclusion (%) 12. Unauthorised absences recorded for children in need (as a % of total possible school sessions) 13. Children in need with SEN receiving SEN support (%) 14. Children in need meeting expected KS2 mathematics level (%) 15. Children in need meeting expected KS2 writing level (%)

		<p>16. Children in need meeting expected KS2 reading level (%)</p> <p>17. Children in need meeting expected KS2 reading, writing and mathematics level (%)</p> <p>18. Children in need meeting expected KS2 grammar, spelling and punctuation level (%)</p> <p>19. Children in need level 4 GCSE pass rate (%)</p> <p>20. Children in need level 5 GCSE pass rate (%)</p> <p>21. Repeat child protection plans (%)</p> <p>22. Child protection plans lasting 2+ years (%)</p>
Children in care (CIC) variables	Procedural	<p>23. Children in care with an up-to-date health assessment (%)</p> <p>24. Children in care with an up-to-date dental check (%)</p>
	Educational outputs and outcomes	<p>25. Children in care 'persistently absent' from education (%)</p> <p>26. Unauthorised absences recorded for children in care (as a % of total possible school sessions)</p> <p>27. Children in care meeting expected KS2 reading level (%)</p> <p>28. Children in care average KS2 progress in writing</p> <p>29. Children in care average KS2 progress in mathematics</p> <p>30. Children in care level 5 GCSE pass rate (%)</p> <p>31. Children in care average Attainment 8 scores</p> <p>32. Children in care average Progress 8 scores</p>
	Health outputs and outcomes	<p>33. Children in care with a substance abuse problem (%)</p> <p>34. Children in care with a substance abuse problem receiving an intervention (%)</p> <p>35. Children in care average SDQ scores</p> <p>36. Children in care with concerning SDQ scores (%)</p> <p>37. Children in care with up-to-date immunisations (%)</p>
	Other outputs and outcomes	<p>38. Children in care placed 20+ miles from home (%)</p> <p>39. Children in care with a missing episode (%)</p> <p>40. Children in care with a criminal conviction (%)</p> <p>41. Children in care with 3+ placement moves in one year (%)</p>

		42. Children in care with a stable placement (%)
Leaving care variables	Outputs and outcomes	43. Young adults (19-21) with care experience who are not in education, employment or training (%) 44. Young adults with care experience living in unsuitable accommodation (%) 45. Young adults with care experience in custody (%)

Statistical Analysis

T-tests and Analysis of Variance (ANOVA) techniques were used to compare differences between the low and high Ofsted categories in relation to procedural, outcomes and population variables. Multivariable ordinal logistic regression modelling was used to evaluate the association between Ofsted ratings and population, procedural and outcome variables. Levels of deprivation and spend per child were used as covariates in all analyses. Child in need (CIN) and child in care (CIC) variables were also included in the final multivariable model. A p-value below 0.05 was considered statistically significant.

Findings

We first analysed the associations between overall ratings and sub-category ratings (for children in need of protection, children in care and leadership). Results revealed very strong and statistically significant correlations. As a result, for the rest of the analysis we used only the authority's overall rating, not the sub-categories.

Associations between Ofsted ratings and key variables

To test the relationships between Ofsted ratings and CIN, CIC and leaving care outcome variables, Spearman's rho non-parametric test were performed. We found one association – for children in need persistently absent from education (Spearman's rho= .29, p=.048).

Spearman's rho non-parametric correlational analyses were also performed in relation to Ofsted ratings and CIN and CIC procedural variables. Statistically significant correlations were identified between Ofsted ratings and two of these variables –proportion of assessments overdue (Spearman's rho=-.347, p=.018), and proportion of CIC with up-to-date dental checks (Spearman's rho=-.364, p=.012). Authorities with better Ofsted ratings had fewer assessments overdue and more CIC with up-to-date dental checks.

For population variables, we found statistically significant correlations between Ofsted ratings and referral rates, rates of CIN and rates of CIC. The higher the Ofsted rating, the fewer referrals, fewer CIN and fewer CIC as a proportion of the total population.

Table 4: Correlations between overall Ofsted ratings and four population variables

	Ofsted overall judgement	Referral rate	CIN rate	CP rate	CIC rate
Referral rate	-.344*		.690**	.618**	.587**
CIN rate	-.403**	.690**		.572**	.554**
CP rate	-.264	.618**	.572**		.714**
CIC rate	-.394**	.587**	.554**	.714**	

Finally, Spearman’s rho non-parametric correlational analyses were performed in relation to Ofsted ratings, spending variables and levels of deprivation. We identified statistically significant negative correlations between Ofsted ratings and deprivation rank (Spearman’s rho= -.421, p=004), Ofsted ratings and total spend per child (Spearman’s rho=-.372, p=.010) and deprivation and spend per child (Spearman’s rho=.417, p=.004). These associations denote that the higher the Ofsted rating, the lower the deprivation rank and the higher the spend per child.

Comparison of means between Ofsted categories

We then calculated the overall mean for a selection of key variables and our two Ofsted categories. We found that authorities in the high Ofsted category demonstrate improved performance in relation to some variables – but not all of them. Authorities with good or outstanding ratings did better in relation to:

- Referral rates
- Repeat referral rates
- CIN rates
- CP plan rates
- CIC rates
- Overdue assessments
- Overdue CP review meetings
- CIC with up-to-date health assessments and dental check-ups

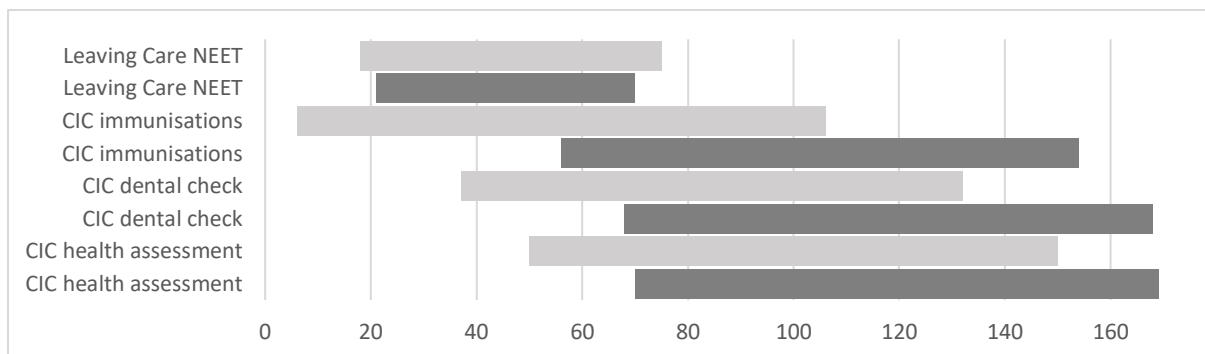
- CIC with interventions for substance abuse problems
- CIC with up-to-date immunisations

On the other hand, authorities with inadequate or requires improvement ratings did better in relation to:

- CIC persistently absent from education
- CP plans lasting two years or more
- CIC GCSE pass rates and educational attainment scores
- CIC with a missing episode
- CIC with a criminal conviction
- CIC with concerning SDQ scores.

We also looked at the range of performance in relation to the high and low Ofsted categories (Figures 2, 3, 4 and 5). These ranges show that even where authorities with high Ofsted ratings outperform authorities with low ratings, the overlap in performance is often substantial, and there is rarely if ever a neat distinction between them. The variables are presented in four different figures because the ranges are too varied to include all of them together.

Figure 2: The range of performance between authorities with high Ofsted ratings (dark grey) and low Ofsted ratings (light grey).



* NEET = not in education, employment or training

Figure 3: The range of performance between authorities with high Ofsted ratings (dark grey) and low Ofsted ratings (light grey).

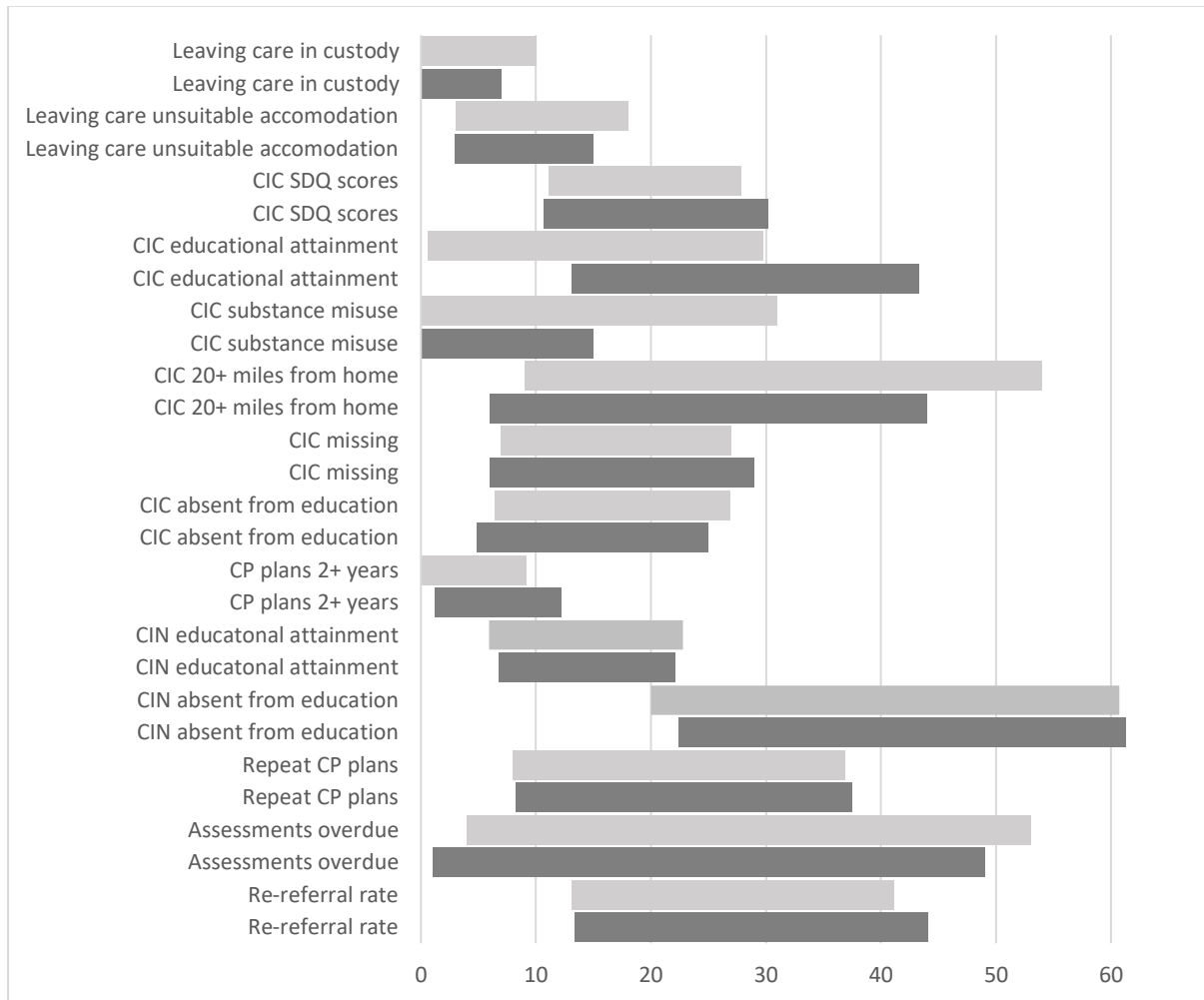


Figure 4: The range of performance between authorities with high Ofsted ratings (dark grey) and low Ofsted ratings (light grey).

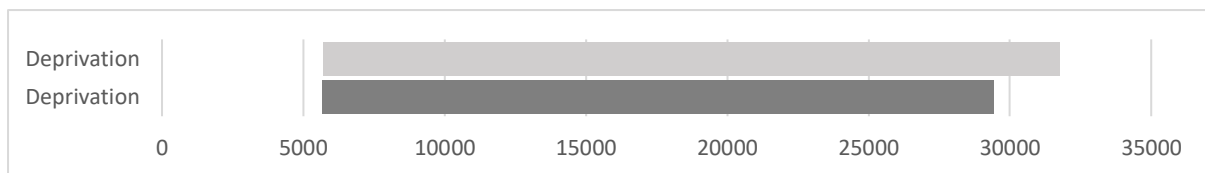
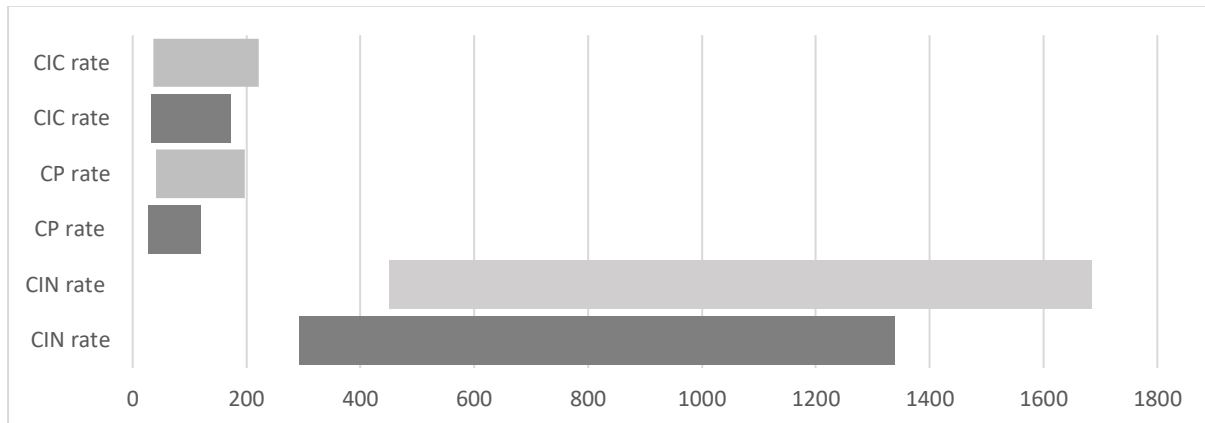


Figure 5: The range of performance between authorities with high Ofsted ratings (dark grey) and low Ofsted ratings (light grey).



Spend per child

Finally, we compared spending rates per child between the Ofsted categories. A significant difference was found for mean total spend per child between the low Ofsted category and the high Ofsted category ($t(45) = 2.26, p = .029$) indicating that the mean spend in the inadequate and requires improvement category was significantly higher than in the good and outstanding category ($M = 829.38, SD = 173.34$ and $M = 714.77, SD = 174.35$, respectively). There were no significant differences for other spending categories ($p > .05$).

Analysis of variance (ANOVA) tests with a Bonferroni correction were carried out to examine differences for total spend per child by level of deprivation. Significant differences were only found within the high Ofsted category ($F(2, 22) = 4.19, p = .03$), indicating that the mean total spend per child for low deprivation areas ($M = 609.30, SD = 97.78$) was significantly lower than the mean spend per child for the medium and the high deprivation areas ($M = 790.65, SD = 159.24$ and $M = 804.33, SD = 227.96$ respectively).

Associations with Ofsted categories

Multivariate logistic regression analyses were conducted to test the association between all the variables we considered and membership of the Ofsted categories through a stepwise process. Only the significant predictors from the univariate regression analyses were included in the final model, that is level of deprivation, up to date dental health checks, assessments overdue and spend per child. The final model included three variables which significantly predicted Ofsted categories ($\chi^2 = 18.815, df = 3, p < .001$) and explained 45.0% of the variance (Nagelkerke $r^2 = .450$). Only the level of deprivation ($B = -.091, Exp b = .913, p = .006$) and up to date dental health checks ($B = .092, Exp b = 1.97, p = .047$) were found to be

significant predictors; though proportion of assessments overdue was not significant at the $\alpha = .05$ level ($B = -.074$, $\text{Exp } b = .929$, $p = .051$), it was very close to this level of significance and was retained in the final model due to the exploratory nature of this study and the limited sample size ($n = 47$). Spend per child was not found to be associated with Ofsted category judgements ($p > .05$), controlling for any other predictor. When controlling for other predictors in the model, the odds of a local authority belonging to the good and outstanding category *decrease* by 8.7% for a one-step change in deprivation rank. In other words, the more deprived the authority, the less likely it is to have received a good or outstanding rating from Ofsted (by nearly 9% for every increase in deprivation ranking).

Table 5. Multivariate model predicting Ofsted category judgement

	$\chi^2(\text{df})$	B	SE(B)	Wald (t-test)	OR	CI(95%)	P-value
Full model	18.518 (3)	-	-	-	-	-	.000***
Predictors							
Deprivation levels		-.091	.033	7.543	.913	(.856, .974)	.006**
Up to date dental health checks	-	.092	.046	3.962	1.097	(1.097, 1.001)	.047*
Assessments Overdue	-	-.059	.038	3.805	.929	(.929, .863)	.051

*. Correlation is significant at the 0.05 level (2

tail).
**. Correlation is significant at the 0.01 level (2-tailed).

Summary and addressing the research questions

In summary, the less deprived the authority and the more CIC have up-to-date dental health checks, the higher the chances of a good or outstanding Ofsted rating (and vice versa).

1. How do Ofsted ratings of Children's Services relate to outcomes for children in need, children in care and young adults with care experience?

Authorities with high Ofsted ratings (good or outstanding) had higher proportions of children in need 'persistently absent' from education. We found no other statistically significant differences.

2. How do Ofsted ratings of Children's Services relate to procedural compliance?

Authorities with high Ofsted ratings (good or outstanding) had lower proportions of overdue assessments and more children in care with up-to-date dental checks.

3. How do Ofsted ratings of Children's Services relate to patterns of spending, deprivation and characteristics of the local population?

Authorities with high Ofsted ratings had lower levels of deprivation and lower overall spending rates per child. They also have lower referral rates, lower rates of children in need and lower rates of children in care. Overall, of the forty-five variables we considered, the authority's deprivation level was the best single predictor of their Ofsted rating.

Strengths and Limitations

The key strength of our study is the use of Department for Education data, which has undergone data normalisation procedures prior to publication.

As with any study, there are a number of limitations. Much of the publicly available data in relation to Children's Services relates to procedural compliance and outputs, rather than outcomes per se. And we had to use spending and deprivation data that may not be truly representative of local contexts. For example, there exists within-authority variation in relation to deprivation. There are also other variables we might have included, such as those related to workforce. The reason we did not include these is partly practical (we can only analyse so much data in one study) and partly to avoid repetition (because these data have been analysed in relation to inspection outcomes before; Hood et al 2016). We were also limited by the relatively small sample size of authorities rated outstanding and inadequate. With larger numbers of such authorities, we would have been able to conduct our analysis without grouping the authorities into two categories.

Discussion

The overall message from our findings is the lack of a consistent association between the variables we included and Ofsted ratings. There is not much that reliably distinguishes between authorities with different Ofsted ratings. In relation to some of the variables, good and outstanding authorities outperform the others, yet the reverse is also true. This is not a situation unique to children's social care. In relation to acute hospitals and GP surgeries in

England, inspection ratings and performance indicators are similarly not well correlated. As noted by Allen et al (2019, 2020), the “*poor power of these indicators to predict subsequent inspection ratings may call into question the validity and reliability of the indicators, inspection ratings, or both*” (p. 55). It is also clear that the extent of outcomes data collected nationally in relation to children’s services, compared with procedural or output data, is very limited. This raises questions about why there is not a clearer association with the variables we included and (or) about the type of performance data collected nationally from local authorities.

How might Ofsted inspectors use this kind of data?

Consider just one of the variables we included – number 41, the proportion of children in care with three or more placement moves in one year (Table 3). We found no consistent difference in relation to this variable between authorities rated better or worse by Ofsted. But what would an Ofsted inspector make of this variable in relation to a particular authority they were inspecting? As Ofsted is now focused in their inspections on the quality of practice, family experiences and outcomes, they would not assume that a high proportion of children with three or more placement moves necessarily indicates a problem. They might well, however, decide to investigate further as a key line of enquiry. They might speak to a number of the social workers involved and find out more about some of the individual cases. They would almost certainly speak to senior managers to see if they were aware of the issue, to ask what they thought was happening and to test the effectiveness of any plans already in place to reduce the number of placement moves. Based on Ferguson et al’s study (Ofsted, 2019a), they would probably not speak to any of the actual children involved. Nonetheless, while they would use such data to inform their inspection methodology, they would not make simplistic judgements based on data alone. What they might find is that some local authorities are more able to give a confident explanation as to why a high proportion of their children are experiencing frequent placement moves, related to the needs of the child and desire to get the best outcomes, while others might explain instead that they have problems recruiting enough foster carers, or about budget cuts forcing children to move from more expensive to less expensive residential placements. These explanations might help to inform – but not determine – the inspection judgement about the quality of practice in each area and offer a richer explanation of practice than would a simple reliance upon the data. (We say ‘not determine’ because a good explanation alone cannot justify poor practice, while good

practice remains so even without a good explanation, albeit inspectors may not always identify it as such without assistance from the local authority.)

The difficulty of evaluating social work practice

By aiming to focus on the quality of practice, Ofsted set themselves a very challenging task. There is limited empirical research on the quality of practice in child and family social work, and only recently have researchers begun to use ‘close to practice’ methods such as ethnographic observation (Ferguson, 2017). As Ofsted note themselves, they do not have the resources to observe social work practice directly. They are reliant on obtaining the views of social workers, managers and senior leaders via interview, as well as gathering evidence from written records. Even assuming that you could reliably evaluate the quality of social work practice from talking and reading about it, the link between what social workers do and ‘outcomes’ for families is not straight-forward (Forrester, 2017). Even when social workers demonstrate advanced practice skills, the influence of these skills on outcomes is relatively modest and mediated by many other factors (Forrester et al, 2019). Ofsted also encounter the same challenge that social workers face when assessing parenting – they can only do so at one point in time, and so it may be the case that any current problems identified are the result of previous patterns of behaviour (or practice) that are no longer evident. Ferguson’s work (2016a, 2016b, 2016c and 2017) has gone a long way to reveal the complexity of what child and family social workers actually do with families, out of which are revealed the complex relationships that must exist between ‘input’, ‘outputs’ and ‘outcomes’ in this context. It is equally true that the relationship between social work practice and outcomes is mediated by more than the behaviour of the worker. Walsh et al (2019) show how responses to family problems are influenced by social workers’ conceptions and definitions of ‘family’, and how these conceptions and definitions are in turn informed by wider organisational factors and social policies. Furthermore, the interaction between many families and children’s services takes place within the context of poverty and deprivation, and there are clear limits to how effective a service can be if it works primarily with individual families, using individual solutions, without being able to take account of structural factors (see Featherstone et al, 2018). Given these complexities, the lack of a clear association between Ofsted ratings and the variables we considered is not so surprising.

The difficulty of evaluating ‘outcomes’

Or perhaps we have simply been looking at the wrong things? Many of the variables in our study were included because the data were available, rather than because of a broad agreement in the field that these things matter. One of the key differences (and there are many) between schools and children's services is that of measurable outcomes. In children's services, it is rather easy to measure outputs – the number of re-referrals, the proportion of timely assessments, the average length of child protection plans and so forth. It is much less easy to measure the kinds of outcomes that social work is often focused upon, such as well-being, civic engagement, and the protection of human rights.

Social work in England, and elsewhere, provides a good case example of the unintended consequences that can occur when attempts are made to improve services by finding more and more 'innovative' ways of measuring what they are doing (Broadhurst et al, 2010). Even if we could agree on a set of meaningful outcomes to measure, there are other aspects of service quality that would remain important - such as whether families feel listened to, respected and included in decision-making (irrespective of whether this improved their 'outcomes'). There is little within the currently available data that directly reflects these qualities, or those of assessment and decision-making, or the protection and promotion of human rights and civic participation, the latter of which some have argued is *the* key purpose of social services (Featherstone, Gupta and Mills, 2018). It remains entirely possible (perhaps even probable) that services rated 'outstanding' are more likely to provide respectful services and to include parents and children effectively in decision-making, for example, than inadequate services - and if so, this would be a very important difference indeed.

Nevertheless, it is hard to argue that *none* of the variables we considered are important. Whether children in care have substance misuse problems, or have good educational attainment, or whether adults with care experience are living in suitable accommodation – these things are important. Is it unreasonable to expect that 'outstanding' services should help more adults with care experience into education, employment or training compared with inadequate services?

Conclusion

According to Ofsted's own ratings, there has been much improvement in the delivery of children's services in England over the past decade, as more authorities receive good and outstanding ratings than before. And powerful arms of the state, such as local authorities, do need regulation (albeit regulation and inspection are not the same thing). Senior leaders cannot always assume they know their own organisations well (Diaz and Aylward, 2018) and

we need mechanisms in place to ensure underperforming services can be identified, supported and where necessary challenged to improve.

At the outset of our study, we anticipated that authorities with good and outstanding ratings would outperform the others, at least in relation to a significant number of the variables we considered. For example, if we had found that good Ofsted ratings were associated with more care leavers living in suitable accommodation and in work or training, better educational results for children ‘in need’ and those in care and fewer children in care with substance misuse problems and criminal convictions, that would have made a powerful argument for the validity of inspection ratings. But this was not what we found. Even the limited outcome measures we were able to include did not correlate with Ofsted’s judgements in a consistent way.

The single best predictor variable was deprivation, something we found previously when looking at Ofsted’s previous inspection framework (Authors Own) and by other researchers in the same field (e.g. Bywaters et al, 2017). Ironically, this association may also be seen as one of the best indicators that Ofsted’s ratings are valid, to the extent we are prepared to accept that deprivation does have a significant impact on the quality of service provided and on family outcomes. If we accept these propositions, we must also question whether a focus solely on improving the quality of practice is justified, as opposed to a wider focus on reducing deprivation – and certainly whether it is reasonable for Ofsted to continue inspecting services without taking the wider socioeconomic circumstances of the authority into account. In reality, it is likely that the quality of practice *and* the authority’s wider socioeconomic circumstances are important mediators of family outcomes (Wijedasa et al, 2018). For the sector as a whole, including those tasked with the critical role of inspection, it is long overdue that this reality was recognised.

Finally, in Ofsted’s (2019b) framework for inspectors, the word ‘outcome’ only appears fourteen times (the document in total is 26,701 words in length). Inspectors are asked to investigate what “*leaders know about practice and outcomes*” (p. 12) and to judge whether practice is “*demonstrably leading to improved outcomes*” (p. 43). Yet in order to ‘measure’ outcomes, it is a prerequisite that they are first defined. The establishment of an agreed outcomes framework would be a significant development for the sector.

References

- Authors Own
- Allen, T., Walshe, K., Proudlove, N. and Sutton, M. (2020) Using quality indicators to predict inspection ratings: Cross-sectional study of general practices in England. *British Journal of General Practice*, 70(690), pp. E55 – E63.
- Allen, T., Walshe, K., Proudlove, N. and Sutton, M. (2019) Do performance indicators predict regulator ratings of healthcare providers? Cross-sectional study of acute hospitals in England. *International Journal for Quality in Health Care*, doi: [10.1093/intqhc/mzz101](https://doi.org/10.1093/intqhc/mzz101)
- Bilson, A. and Martin, K. (2017) Referrals and child protection in England: one in five children referred to Children’s Services and one in nineteen investigated before the age of five. *British Journal of Social Work*, 47(3), pp. 793 – 811.
- Brilliant, E. (1986) Social work leadership: A missing ingredient? *Social Work*, 31(5), pp. 325 – 331.
- Broadhurst, K., Wastell, D., White, S., Hall, C., Peckover, S., Thompson, K., Pithouse, A. and Davey, D. (2010) Performing ‘initial assessment’: Identifying the latent conditions for error at the front door of local authority children’s services. *British Journal of Social Work*, 40(2), pp. 352 – 370.
- Bywaters, P., Brady, G., Sparks, T., Bos, E., Bunting, L., Daniel, B., Featherstone, B., Morris, K. and Scourfield, J. (2015) Exploring inequities in child welfare and child protection services: Explaining the ‘inverse intervention law’. *Children and Youth Services Review*, 57(1), pp. 98 – 105.
- Bywaters, P., Webb, C. and Sparks, T. (2017) Ofsted ratings do reflect local authority deprivation and spending. Available at: <https://www.communitycare.co.uk/2017/01/18/ofsted-ratings-reflect-local-authority-deprivation-spending/> [Accessed 7th November 2019]
- Bywaters, P. and Webb, C. (2018) There is clear evidence that links deprivation, expenditure and quality in Children’s Services. Available at: <https://www.communitycare.co.uk/2018/02/07/clear-evidence-links-deprivation-expenditure-quality-childrens-services/> [Accessed 21st May 2019]
- Campbell, J. and Husbands, C. (2000) ‘On the reliability of OFSTED inspection of initial teacher training: A case study’, *British Educational Research Journal*, 26(1), pp. 39–48.

- Diaz, C. and Aylward, T. (2018) A study on senior managers' views of participation in one local authority... a case of wilful blindness? *British Journal of Social Work*, doi.org/10.1093/bjsw/bcy101
- Featherstone, B., Gupta, A. and Mills, S. (2018) The role of the social worker in adoption – ethics and human rights: an enquiry. Available at: <https://bettercarenetwork.org/sites/default/files/The%20Role%20of%20the%20Social%20Worker%20in%20Adoption%20Enquiry.pdf> [Accessed 7th November 2019]
- Featherstone, B., Gupta, A., Morris, K. and Warner, J. (2018) Let's stop feeding the risk monster: towards a 'social model' of child protection. *Families, Relationships and Societies*, 7(1), pp. 7 – 22.
- Ferguson, H. (2016a) Making home visits: Creativity and the embodied practices of home visiting in social work and child protection. *Qualitative Social Work*, 17(1), pp. 65 – 80.
- Ferguson, H. (2016b) What social workers do in performing child protection work: evidence from research into face-to-face practice. *Child and Family Social Work*, 21(3), pp. 283 - 294.
- Ferguson, H. (2016c) Researching social work practice close up: Using ethnographic and mobile methods to understand encounters between social workers, children and families. *British Journal of Social Work*, 46(1), pp. 153 - 168.
- Ferguson, H. (2017) How children become invisible in child protection work: Evidence from day-to-day social work practice. *British Journal of Social Work*, 47(4), pp. 1007 - 1023.
- Forrester, D. (2017) Outcomes in Children's Social Care. *Journal of Children's Services*, 12(2-3), pp. 144 – 157.
- Forrester (2019) What is the point of child and family social work? Available at: <http://sites.cardiff.ac.uk/cascade/2019/05/15/what-is-the-point-of-child-and-family-social-work/> [Accessed 21st May 2019]
- Forrester, D., Westlake, D., Killian, M., Antonopolou, V., Thurnham, M., Thomas, R., Waits, C., Whittaker, C. and Hutchinson, D. (2019) What is the relationship between worker skills and outcomes for families in child and family social work? *British Journal of Social Work*, bcy126, <https://doi.org/10.1093/bjsw/bcy126>

- *The Guardian* (2015) Is it time to change the way Children’s Services are inspected? Available at: <https://www.theguardian.com/social-care-network/2015/oct/14/is-it-time-to-change-the-way-childrens-services-are-inspected> [Accessed 20th May 2019]
- Höjer, I. (2011) Parents with children in foster care – how do they perceive their contact with social workers? *Practice*, 23(2), pp. 111 – 123.
- Hood, R., Grant, R., Jones, R., and Goldacre, A. (2016) A study of performance indicators and Ofsted ratings in English child protection services. *Children and Youth Services Review*, 67, pp. 50 – 56.
- Hood, R., Nilsson, D. and Habibi, R. (2018) An analysis of Ofsted inspection reports for children’s social care services in England. *Child and Family Social Work*, 24(2), pp. 227 – 237.
- Impower (2015) ‘A brave new world: Is inspection improving Children’s Services?’, available online at: <https://www.impower.co.uk/reports/a-brave-new-world-is-inspection-improving-childrens-services> [Accessed 20th May 2019]
- Lawler, J. (2007) Leadership in social work: A case of caveat emptor? *British Journal of Social Work*, 37(1), pp. 123 – 141.
- Jones, R. (2015) Ofsted doesn’t recognise the practice chaos and professional carnage it leaves in its wake. Available at: <https://www.communitycare.co.uk/2015/05/07/ofsted-doesnt-recognise-practice-chaos-professional-carnage-leaves-wake/> [Accessed 20th May 2019]
- Matthews, A. (2018) Social worker on inspections: The drive to compliance in the run up to Ofsted visit was striking. Available at: <https://www.communitycare.co.uk/2018/03/08/social-worker-inspections-drive-compliance-run-ofsted-visit-striking/> [Accessed 20th May 2019]
- National Audit Office (2018) Ofsted’s inspections of schools. Available at: <https://www.nao.org.uk/report/ofsteds-inspection-of-schools/> [Accessed 20th May 2019]
- Ofsted (2014) The new Ofsted framework for the inspection of Children’s Services and for reviews of Local Safeguarding Children Boards: an evaluation. Available at: <https://dera.ioe.ac.uk/20253/1/The%20new%20Ofsted%20framework%20for%20the%20inspection%20of%20children%E2%84%A2s%20services%20and%20for%20reviews%20of%20LSCBs%20an%20evaluation.pdf> [Accessed 21st May 2019]

- Ofsted (2017) The Annual Report of Her Majesty’s Chief Inspector of Education, Children’s Services and Skills 2016/17. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/666871/Ofsted_Annual_Report_2016-17_Accessible.pdf [Accessed 21st May 2019]
- Ofsted (2019a) Inspection of local authority’s children’s services framework implementation review. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/834443/ILACS_framework_implementation_review.pdf [Accessed 7th November 2019]
- Ofsted (2019b) Framework, evaluation criteria and inspector guidance for the inspection of local authority children’s services. Available at: <https://www.gov.uk/government/publications/inspecting-local-authority-childrens-services-from-2018> [Accessed 7th May 2020]
- Richmond, T. (2019) Requires improvement: a new role for Ofsted and school inspections. Available at: <https://www.edsk.org/wp-content/uploads/2019/04/Requires-Improvement.pdf> [Accessed 20th May 2019]
- Rogers, T. (2017) Ofsted has done more to damage education in the UK than anything else. Available at: <https://www.tes.com/news/ofsted-has-done-more-damage-education-uk-anything-else> [Accessed 20th May 2019]
- Rosenthal, L. (2004) ‘Do school inspections improve school quality? Ofsted inspections and school examination results in the UK’, *Economics of Education Review*, 23(2), pp. 143–51.
- Schooling, E. (2017a) *A new inspection system*. Available at: <https://socialcareinspection.blog.gov.uk/2017/11/29/a-new-inspection-system/> [Accessed 20th May 2019]
- Schooling, E. (2017b) *Process still matters in social work – but not as much as outcomes*. Available at: <https://socialcareinspection.blog.gov.uk/2017/11/29/a-new-inspection-system/> [Accessed 20th May 2019]
- Shaw, I., Newton, P., Aitkin, M. and Darnell, R. (2003) ‘Do OSFTED inspections of Secondary Schools Make a Difference to GCSE Results? ’, *British Educational Research Journal*, 29(1), pp. 63–75.

- Shoemith, S. (2016) *Learning from Baby P: The politics of blame, fear and denial*. London, Jessica Kingsley.
- Stevenson, L. (2016) Ofsted gives first outstanding grades for Children's Services to two London councils. Available at:
<https://www.communitycare.co.uk/2016/03/29/ofsted-gives-first-outstanding-grades-childrens-services-two-london-councils/> [Accessed 20th May 2019]
- Walsh, J., White, S., Morris, K. and Doherty, P. (2019) How do you solve a problem like Maria? Family complexity and institutional complications in UK social work. *European Journal of Social Work*, 22(6), pp. 1050 – 1061.
- Wijedesa, D., Warner, N. and Scourfield, J. (2018) Exploratory analyses of the rates of children looked after in English local authorities (2012-2017). London, What Works for Children's Social Care.