

Accepted Manuscript

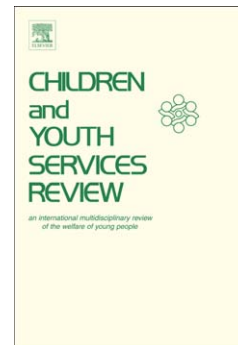
A study of performance indicators and Ofsted ratings in English child protection services

Rick Hood, Robert Grant, Ray Jones, Allie Goldacre

PII: S0190-7409(16)30167-0
DOI: doi: [10.1016/j.chilyouth.2016.05.022](https://doi.org/10.1016/j.chilyouth.2016.05.022)
Reference: CYSR 2982

To appear in: *Children and Youth Services Review*

Received date: 5 April 2016
Revised date: 29 May 2016
Accepted date: 29 May 2016



Please cite this article as: Hood, R., Grant, R., Jones, R. & Goldacre, A., A study of performance indicators and Ofsted ratings in English child protection services, *Children and Youth Services Review* (2016), doi: [10.1016/j.chilyouth.2016.05.022](https://doi.org/10.1016/j.chilyouth.2016.05.022)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Title: A study of performance indicators and Ofsted ratings in English child protection services

Author names and affiliations.

1. Rick Hood, Kingston University and St Georges, University of London
School of Social Work
Faculty of Health, Social Care and Education
Kingston University and St George's, University of London
Kenry House; Kingston Hill
Kingston-upon-Thames KT2 7LB
Email: Rick.Hood@sgul.kingston.ac.uk
2. Robert Grant, Kingston University and St Georges, University of London
Faculty of Health, Social Care and Education
St Georges, University of London
2nd Floor, Grosvenor Wing
Cranmer Terrace
London SW17 0RE
Email: Robert.Grant@sgul.kingston.ac.uk
3. Ray Jones, Kingston University and St Georges, University of London
School of Social Work
Faculty of Health, Social Care and Education
Kingston University and St George's, University of London
Kenry House; Kingston Hill
Kingston-upon-Thames KT2 7LB
Email: ray.jones1@hotmail.co.uk
4. Allie Goldacre, Kingston University and St Georges, University of London
School of Social Work
Faculty of Health, Social Care and Education
Kingston University and St George's, University of London
Kenry House; Kingston Hill
Kingston-upon-Thames KT2 7LB
Email: alliegoldacre@gmail.com

Corresponding author: Rick Hood (as above)

Abstract

This paper presents new findings from a study of performance measures for children in need and child protection services in England. National datasets and census returns from 152 local authorities over a 13-year period were combined in order to analyse trends and correlations in quality indicators. The study also explored the relationship between these measures and inspection ratings from the Office for Standards in Education, Children's Services and Skills (Ofsted), with a particular focus on services rated as inadequate. The available quality measures mainly focused on the timeliness of work processes, but these did not seem to affect outcomes in the form of re-referral rates. However, re-referrals were higher in local authorities with a tendency to close cases quickly and in those with high rates of agency workers. A small number of indicators were able to predict an inadequate Ofsted rating in 2012 and 2013. Changes in performance measures in the year following an inadequate Ofsted rating may suggest greater use of protective interventions compared with similarly performing local authorities. Implications are considered for performance measurement, management and inspection in the field of child protection.

Keywords

Child protection, child welfare, Ofsted, performance indicators, management, inspection

Introduction

Performance management, the process through which an organisation's managers evaluate their employees' work and distribute rewards in order to achieve strategic goals, has been a feature of UK public services since the early 1980s (Osbourne et al., 1995; Baars, Evers, Arntz, & van Merode, 2010). It gathered pace during the New Labour administration of 1997-2010, during which the use of performance-related indicators, benchmarks, targets, and incentives in the public sector become widespread (Propper and Wilson, 2003). The increasing focus on performance can be connected to the New Public Management (NPM) approach, which over the same period has sought to make the professional bureaucracies originally established by the post-war welfare state more 'business-like', and in the process make professional groups more accountable to service users and taxpayers (Hood, 1991; Cochrane, 2000). The 'doctrinal components' of NPM include a commitment to 'explicit standards and measures of performance', and 'greater emphasis on output controls' (Hood, 1991: 4). Over time, these components have become associated with a regime of internal audit and external inspection, reinforced by IT-based workflow systems and the threat of sanctions for non-compliance (Bevan and Hood, 2006).

These ideas and reforms have greatly influenced the design and delivery of child protection services over recent decades. Frontline practice has been transformed by the introduction of electronic workflow systems that not only shape assessment and intervention processes but also gather statistics for managerial and quality assurance purposes (Shaw et al., 2009; Munro 2004). The expansion of audit and performance monitoring has also been accompanied by an increasingly robust approach to external inspection, which in relation to children and families services is currently carried out by the Office for Standards in Education, Children's Services and Skills (Ofsted, 2015a). Such developments have been criticised for focusing on process outputs rather than user outcomes (Munro, 2004; Tilbury, 2004), and shaping child protection into a technocratic exercise that revolves around compliance with procedures and standards (Ayre and Preston-Shoot, 2010). Organisations and professionals have sought to use performance data to reduce the uncertainty inherent in such a complex field, driven by anxiety about making the wrong decision and of public criticism (Lees, Meyer & Rafferty,

2013; Munro, 2010). Accountability and performance have therefore become increasingly associated with risk and risk management (Hood, 2015).

Performance-based accountability

‘Performance-based accountability’ is a framework for conceptualising the distinction between processes and outcomes in performance management (Friedman, 1997; 2001). Processes in relation to child welfare include inputs and resources, such as numbers of staff and caseloads, as well as the time taken to complete important pieces of work, such as needs assessments. Outcomes relate to the effects of intervention; taking the example of children subject to child protection (CP) plans, outcomes might include the number of children who are stepped down from CP plans within a certain time frame, or the proportion of plans made for children who have already had this service in the past. These distinctions can be used to create a typology of indicators based on how they are measuring performance, a version of which is represented in Figure 1 below:

Figure 1: Performance-based accountability (adapted from Friedman, 1997: 4-5)

	Quantity	Quality
Effort	How much did we do?	How well did we do it?
Effect	Did anything change?	Was it change for the better?

Figure 1 distinguishes not only between ‘effort’ (i.e. inputs) and ‘effect’ (i.e. outcomes), but also between quantity and quality. Friedman (1997) notes that the most significant data are provided by ‘quality’ measures, particularly those in the bottom right quadrant that indicate the quality of ‘change for the better’ produced by the service. Unfortunately, these are also

the most difficult measures to obtain, partly because the nature of outcomes often takes a long time to emerge. As a result, a lot of the quality measures organisations tend to collect a tend to be skewed towards effort ('how well did we do it') rather than effect ('change for the better'). This is illustrated below in Table 1, which uses the matrix shown in Figure 1 to categorise the main quality measures for children's social care that are collected and made publically available by the Department of Education (DfE, 2015a; 2015b) and Ofsted (2015b).

Table 1. Common quality indicators for English children's social care services

QUALITY OF EFFORT: How well did we do it?
<ul style="list-style-type: none"> • Numbers of CIN per social worker • Social work vacancies • Turnover of social workers • Percentage of agency workers¹ • Core assessments completed within 35 days • Continuous Single Assessments completed within 45 days • Assessments as a percentage of referrals • CP conference held within 15 days of Section 47 • CPP cease times (<3 months, 3-5 months, 6-11 months, 1-2 years, 2+ years) • Initial Assessments completed within 7/10 days • CP Plans review held within 6 months • CIN Cease times (<3 months, 3-5 months, 6-11 months, 1-2 years, 2+ years) • Ofsted ratings

QUALITY OF EFFECT

Was it change for the better?

- CP Plan where children had prior CP Plan
- Referrals within 12 months of a prior referral
- Ofsted ratings

Note

¹ In England, most statutory social workers are public sector workers, i.e. they are employed on a permanent basis by local authorities. However, a minority of vacancies may be filled on a short-term basis through private social work agencies, and these employees are often termed ‘locum’ or ‘agency’ workers.

It will be apparent that Table 1 has only three ‘quality of effect’ measures, and two of these relate to the proportion of re-referrals and re-registrations – i.e. they are a ‘negative’ measure in the sense of highlighting the rate of cases that are not dealt with first time round. In other words, ‘change for the better’ is assumed to involve a reduction in both of these measures. Barth and Reid (2000) point out that child welfare services have tended to ignore what happens to children after their involvement with services ends, although connecting CIN statistics to information on educational achievement from the National Pupil Database has recently been suggested (DfE, 2015c). Ofsted inspections are a wide-ranging review of service delivery in public sector organisations. As such, inspection ratings reflect a range of qualitative and quantitative data including some of the indicators listed above, and have therefore been categorised here as both a quality-of-effort and quality-of-effect measure.

Method

In this study, the following questions were posed:

1. How have indicators of quality varied over time?
2. How do different measures of quality relate to each other, i.e. are changes in one correlated with changes in any of the others?
3. Is there a connection between performance indicators and Ofsted ratings?

4. What happens to indicators after an Ofsted inadequate rating?

The procedures used for gathering and combining the national datasets are described in Hood et al. (2016). A comprehensive set of indicators of local authority child protection services were obtained for the period 2001 – 2014 from the UK Government website, the National Archives online, and the Cafcass website. Indicators were converted to rates per 10,000 population, and two small and unusual authorities, the City of London and the Isles of Scilly, were excluded from the analysis. This paper focusses on the quality indicators set out in the right hand column of Table 1. To address the first research question, national trends were plotted over time. For the second question, significant Spearman's correlations among the indicators were compared for each year in 2009-14. This period had comparable data from the CIN census as well as workforce.

For the third question, a regression model was built aiming to obtain the closest prediction of the binary outcome of having an “inadequate” Ofsted rating in 2012 or 2013, using quality and quantity indicators for 2012 only (the year with best quality data). We used elastic-net regression via the ‘glmnet’ package (version 2.0-2) for R (version 3.2.2), which produces a limited set of the most important indicators. This technique has been shown to be more reliable than stepwise regression (Zou and Hastie, 2005). The cross-validation function within glmnet was used to identify the best model.

For the fourth question, all local authorities (LAs) predicted to get an inadequate rating during 2010/11 – 2012/13 were divided into two groups according to whether they actually received an inadequate rating or not. Those LAs predicted by the elastic net regression as having more than 50% chance of being rated inadequate were classified as Group A. Those LAs who were predicted to be rated inadequate and did in fact receive an inadequate rating were classified as Group B. The two groups can be regarded as false positives and true positives respectively. This comparison sought to exclude the problem of regression to the mean: indicators improving after an unusually bad year simply due to random variation.

Throughout this paper, statistics represent all cases at all authorities, so it should be borne in mind that inference, including p-values, can only be interpreted in terms of future performance. However, we know that the child protection system is a complex one and it is likely to continue to be subject to fundamental policy-driven changes. So, all results should

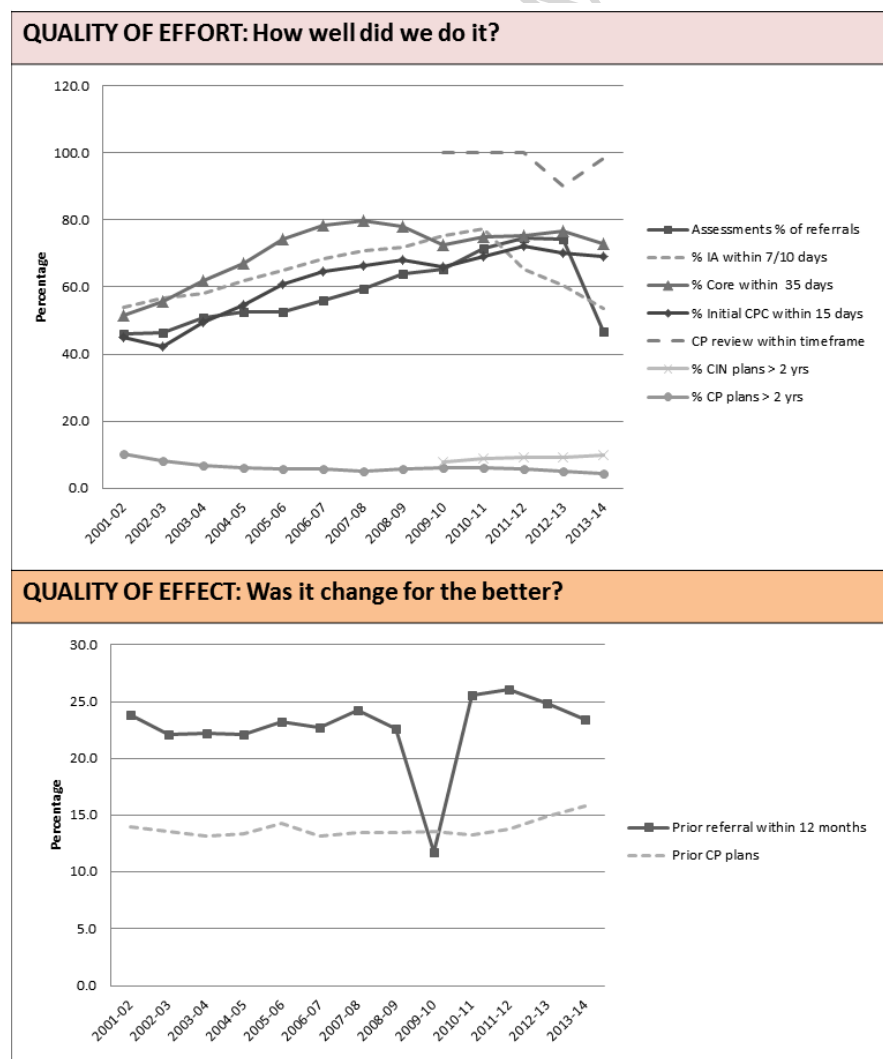
only be regarded as giving statistical insights into the processes at work during this period of time.

Results

Analysis of trends

A longitudinal analysis of quality measures is illustrated in Figure 2, covering a 13 year period from 2001 – 2014. The results have been categorised in line with the performance matrix outlined above (see Table 1) and represent median rates per 10,000 population for 150 English local authorities, i.e. national trends.

Figure 2. Quality measures for English local authorities 2001/02 – 2013/14



The 'quality of effort' indicators in the top half of Figure 2 show some of the work processes already quantified above, such as assessments and case conferences, for which several years' data is available. The indicators track the percentage of these processes achieving set criteria, which following statutory guidelines have tended to revolve around completion within a given timeframe, e.g. the percentage of initial child protection case conferences held within 15 working days of the relevant strategy discussion (DCSF, 2010). Timeliness indicators for holding CP conferences and completing assessments show a steady increase over the period 2001-07 suggesting that local authorities moved towards more timely completion rates in line with statutory guidelines, with a ceiling of about 80% for most indicators. The picture is more mixed thereafter, with a particularly noticeable decline in the timeliness of initial assessment completion from 2009-13.

The other timescale indicators show the percentage of CIN plans and CP plans that lasted for more than two years before the children ceased to be on a plan. These measures are somewhat ambiguous; while it is reasonable to assume that families do not want prolonged involvement with child protection services, it does not automatically follow that longer interventions are necessarily a sign of ineffectiveness and drift. The analysis shows that the proportion of CP plans lasting over 2 years has been falling while the proportion of CIN plans lasting over 2 years has been rising. However, these are relatively small proportions of overall interventions; around 50% of CIN cases cease within 3 months, and most CP plans (around 40%) cease within 6-11 months. Another quality of effort measure shown in Figure 2 is that of assessments completed as a percentage of referrals. This indicator also rose steadily over a 12 year period before falling steeply in 2013. The drop in the last year of data could reflect a delay in achieving full data on the system, or may be a knock-on effect of the introduction of continuous single assessments.

The bottom half of Figure 2 focuses on two quality-of-outcome measures: the percentage of re-referred cases over the past twelve months, as well as the percentage of CP plans made for children who had previously had a CP plan. The most evident feature is the sharp dip in re-referrals occurring in 2009-10. Prior to this the indicator had fluctuated between 22% and 24%; afterwards it rose to 26% before declining subsequently. The unusual volatility may be a statistical outlier resulting from the change to CIN census data in that year. Alternatively, the indicator might be capturing a period during which work was being held in the system

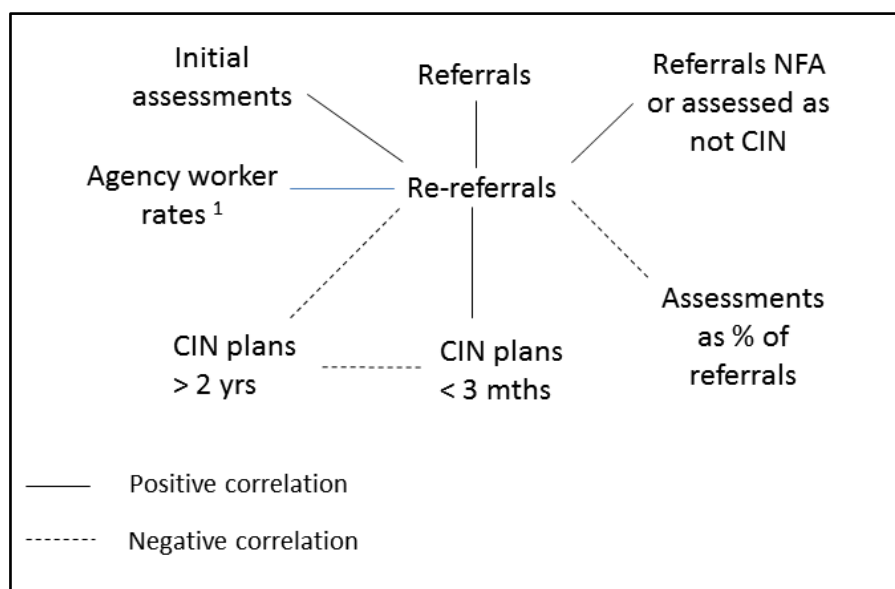
longer than usual before being released. The pattern is not replicated in the equivalent measure for repeat CP plans, which were in slow decline for most of the period before rising steeply from 2010 to 2014.

Analysis of correlations

The results reported here highlight correlations that were significant ($p < 0.05$) across three or more years in the case of CIN indicators, and across two years for workforce indicators (which were only available in 2012-13 and 2013-14). Looking first at ‘quality of effort’ measures, there was unsurprisingly a correlation between timescale indicators for initial and core assessments, i.e. local authorities that completed initial assessments on time also tended to complete core assessments on time. However, these two indicators were not correlated with timescale compliance in relation to CP case conferences. Furthermore, there was no correlation between timescale indicators and any of the quantity measures, such as referrals or numbers of CIN, nor with deprivation. This suggests that local authorities handling higher quantities of work were just as likely to comply with timescales as those with fewer assessments to do.

Moving on to ‘quality of effort’ measures, there were no significant correlations found between ‘prior CP plans’ and other indicators, including with quantity measures for CP interventions. However, a number of correlations were found in relation to ‘re-referrals within 12 months’, as shown below in Figure 3:

Figure 3. Significant correlations with re-referral rates (2009-14)



Note: ¹ data for agency worker rates were only available for 2012-14

Figure 3 shows how outcomes in the form of re-referral rates were linked to other aspects of performance. Re-referrals were naturally correlated with referrals, of which they are a subset. Having a higher percentage of referrals screened out for non-statutory support (not CIN) or no further action (NFA) was associated with more re-referrals. The converse was also true, in that conducting more assessments as a percentage of referrals was negatively correlated with both referrals and re-referrals.

As reported by Hood et al (2016), re-referrals correlated positively with CIN plans ceasing within 3 months, and negatively with CIN plans ceasing over two years. Shorter duration of CIN plans, as well as a greater tendency to screen out referrals rather than assess them, was therefore associated with poorer outcomes in terms of re-referral rates. Finally, an analysis of ‘workforce indicators’ for children’s social workers, i.e. rates for staff turnover, vacancies and agency workers, was undertaken for the two years these data were available. One of these indicators, the percentage of social workers employed via agencies, correlated positively with re-referral rates, suggesting that high rates of agency workers were a negative quality indicator.

Analysis of Ofsted ratings

The best predictive model was achieved with the coefficients listed in Table 2 below. These variables together explained 40% of the deviance (analogous to the R-squared value in linear regression), which represents an encouraging fit to the data in such a complex scenario. The model is dominated by three larger coefficients, shown in bold underlined type. Together they suggest that timely initial assessment is associated with lower risk of an “inadequate” rating, shown by the negative coefficient, while higher rates of agency workers and re-referrals are associated with an increased risk. The accuracy of the model’s predictions showed a sensitivity of 68% (ability to spot true positives) and 94% specificity (ability to spot true negatives). The model’s three components were to some extent congruent with the correlation analysis (see Figure 3), which showed the re-referrals to be a key outcomes indicator associated with agency worker rates. It is also worth noting that performance on

initial assessment timescales, which was not correlated to re-referrals, was in steady decline during the period in question (see Figure 2).

Table 2. Summary of predictor variables for Ofsted ratings in 2012-13

Predictor variable	Coefficient
% of initial assessments taking place within 10 days in 2012	<u>-0.051</u>
Rate of Section 47s during 2012	0.001
% of referrals with a prior referral during 2012	<u>0.028</u>
Social worker vacancy rate 2012	0.004
Social worker agency worker rate 2012	<u>0.076</u>

The model was then used on data from 2010-2012 to compare (A) local authorities who were predicted to be inadequate but were not rated as such, with (B) those who were predicted and then did actually get an inadequate rating. For both groups, indicators in the *first* year they predicted an inadequate rating were compared with the year after that prediction. Group A consisted of 16 local authorities (LAs), eight of which had no inspection over the period in question and therefore did not receive a rating. Of the eight LAs who did receive a rating, one was rated ‘good’ and seven were rated ‘3’, one rating below inadequate, which represented an ‘adequate’ judgement at the time (since 2013 this rating has been re-labelled as ‘requires improvement’). In Group A, the median percentage of initial assessments completed within 10 days was 61%, with an interquartile range (IQR) of 56-65%. The median percentage of agency social workers was 1% (IQR 0-13%) and of re-referrals was 25% (IQR 19-30%). For local authorities in Group B, all of which were inspected and rated inadequate during the

same period, the equivalent figures were 58% for initial assessments (IQR 46-63%), 24% for re-referrals (IQR 20-29%), and 32% for agency workers (IQR 16-34%).

The results of the comparison between Group A and B are shown below in Table 3. Median change is shown as well as the percentage of authorities where the indicator value rose. Whether an increase is desirable or not depends on the indicator, and may be ambiguous in some instances. For example, an average rise in the timeliness of assessments would generally be considered a positive change, whereas an increase in re-referral rates would not. Changes in quantity indicators, such as rates of CP plans, are more open to interpretation.

Table 3. Movement in performance indicators following an 'inadequate' Ofsted rating

Indicator	Movement in indicators from the first year of being <u>predicted</u> inadequate to the following year			
	Group A (n=16) Predicted BUT NOT rated inadequate 2010-12		Group B (n=12) Predicted AND rated inadequate 2010-12	
	<i>median change</i>	<i>% LAs where indicator rises</i>	<i>median change</i>	<i>% LAs where indicator rises</i>
Rates of CIN	-46.6	31	30.2	83
Rates of Referrals	-8.2	44	-0.85	50
Rates of Initial Assessments	-29.75	19	-54.2	27
Initial Assessments < 10 days	4.9	79	4.6	64
Rates of Core Assessments	12.1	67	38.5	100
Core Assessments < 35 Days	3.75	58	15.75	100
Rates of Section 47 inquiries	2.8	56	27.05	58

Rates of CP Conferences	2.7	62	6.2	58
Initial CP Conference < 15 days	3.9	53	0.5	50
Rates of CP plans	3.5	69	8.5	67
Review CP conference on time	-0.05	50	-0.15	50
Rates of Care Proceedings	0.1	56	-0.65	33
Rates of re-referrals	-2.45	31	-1.85	42
Rates of repeat CP plans	-0.25	44	2.1	83
Spending on CP services	18	50	537	75
Rates of CIN per social worker	-1.8	42	-2.2	42
Vacancy Rates	0.65	75	3.05	67
Turnover Rates	1.35	50	0.35	50
Agency Worker Rates	0	58	0.2	50

These were small groups of authorities and so differences between the two groups on the level of individual indicators were not statistically significant. However, comparing the two groups helps to show the differences associated with an inadequate rating, given that any given set of indicators could be expected to vary somewhat from one year to the next. Interestingly, the overall pattern across indicators was that local authorities that had been rated as inadequate by Ofsted seemed to move towards higher levels of child protection intervention and planning (e.g. Section 47 investigations, CP plans, CP conferences) than did the comparison group. This in turn was accompanied by much higher spending on child protection services but little change in caseloads and a slightly higher rate of staff vacancies in services that were rated inadequate. Re-referral rates improved slightly for both groups but there was possibly more of a tendency for 'inadequate' authorities to place children back on CP plans. Again, this was not significant from a statistical point of view but reflected the

overall increase in national figures for this indicator during the period in question (see Figure 2).

Summary of findings

- Quality measures mainly focused on the timeliness of work processes, but these did not seem to affect outcomes in the form of re-referral rates.
- Re-referrals, which are a negative outcomes measure, were higher in local authorities with a tendency to close cases quickly and in those with high rates of agency workers.
- There was a sharp dip in re-referrals in 2009/10, which may have been the result of cases being 'held' in the child protection system in the previous year.
- A small number of indicators were able to predict, with 68% accuracy, an inadequate Ofsted rating in 2012 and 2013. These indicators were: timeliness of initial assessments, rates of re-referrals, and agency worker rates
- Changes in performance measures in the year following an inadequate Ofsted rating seemed to suggest greater use of protective interventions compared with similarly performing local authorities.

Discussion

Performance is an ambiguous concept that combines ideas of functionality (how effectively and efficiently a service meets the purpose for which it was designed), comparability (how it compares to other similar services) and compliance (how it satisfies criteria set out in statutory guidelines and enforced by regulators and inspectors) (Smith, Mossialos & Papanicolas, 2009). The potential pitfall is that indicators might focus on some aspects of performance but not others, and the results presented above do raise some problematic questions in this regard. The most obvious issue is the dearth of outcomes measures. It is rather telling that the only quality-of-effect indicators available – re-referrals and repeat CP plans – actually measure failure rather than success. This not to dismiss the importance of measuring 'failure demand', which has been shown to be a very useful way of highlighting distortions in systems design (Seddon, 2008). Nonetheless it seems perverse that a well-

known proxy for inefficient and fragmented service delivery should become the default outcomes measure for children's social care.

Paucity of information about the quality of outcomes is not unusual in the health and social care sector (Smith, Mossialos & Papanicolas, 2009). However, it does put the onus on process indicators, ideally those with a clear connection to the benefits of intervention. A range of 'quality-of-effort' indicators were explored in this study. Adherence to work completion deadlines, or their decline in relation to initial assessment after 2010-11, seemed to have no effect on outcomes as measured by re-referrals. Other time-based measures were also problematic, with earlier case closure positively correlated with quantity of work as well as re-referral rates, and so apparently geared more towards managing high levels of demand than addressing need more efficiently. The converse finding that *longer* interventions (duration of CIN and CP plans) were associated with better outcomes (lower re-referral rates) rather contradicts the signs of timeliness and efficiency that one might expect to find in high quality services, although it is perhaps less surprising in the context of families with entrenched problems and complex needs (Own Author, 2012). There does seem something strange about a system that sets timeliness and problem-solving at odds with each other, however, and by extension the duration of organisational workflows such as CIN and CP plans seems less meaningful for performance if it does not accurately reflect turnaround times from the family's perspective (Gibson and O'Donovan, 2014).

Another issue highlighted by the findings is the way in which performance measures are affected by events beyond the control of the organization itself. The most obvious example is the effect of public scandals about deaths from child abuse, which have immediate but also longer-term repercussions for child welfare services across the country (Hood et al., 2016; Cafcass, 2012). Indeed, what Sass and Crosbie (2013) call a 'scandal-reform cycle' has shaped these services for decades, not just in England but also in other countries (Lonne, Parton, Thomson & Harries, 2008). During the period covered in this study, a crisis erupted in late 2008 following media coverage of the death of Peter Connelly. Peter was a 17 month old little boy – known initially in the media as 'Baby P' due to court reporting restrictions – who had died at home the previous year following terrible neglect and abuse. He had been known to child protection services at the time of his death, and the ensuing media campaign of vilification directed at those services (Jones, 2014) was associated with a surge in the use of child protection interventions that in England continues to gather pace (Hood et al., 2016).

Movements in performance indicators – such as the dip in re-referral rates reported in 2009/10 – should therefore also be seen in the context of institutional anxiety about risk (Rothstein, Huber and Gaskell, 2008) and its escalation at times of crisis (Cafcass, 2012).

The quality measures discussed up to now allow local authorities to manage their compliance with statutory guidelines, track the timeliness of different elements of provision, and explore trends in service delivery that may be indicative of poor performance. The tendency for these indicators to be subject to mandatory reporting and incorporation into national datasets reflects burgeoning government interest in the possibilities of 'big data' to balance accountability and austerity (HM Government, 2013). In children's social care, as in education, the main context for performance measurement is inspection by Ofsted, which for many local authorities now presents a risk of being judged inadequate or requiring improvement (ADCS, 2015). The issue of whether the inspection regime has become excessively punitive is linked to wider political and strategic considerations (Jones, 2015). In terms of performance management, the main question is whether inspections are an effective mechanism for evaluating and improving the quality of services. There are a number of aspects to this question. First, do ratings reflect quality? Second, do the reports add anything to existing indicators? And third, do inspections help services to improve? All these questions require more detailed qualitative research, but some implications can be drawn from the present study.

First, the findings of this study suggest that ratings (at least in the two years considered here) were congruent with an overall picture of performance shown by a combination of quality indicators around timeliness, workforce stability, and failure demand. Many local authorities whose indicators suggested poor performance were indeed rated as inadequate, and very few local authorities rated as inadequate had statistics that belied their rating. Of course, this predictive ability may partly be due to the role that such quality indicators play in the inspection process itself. However, they may also constitute 'warning signs' of poor performance in the more holistic sense conveyed by an Ofsted judgement.

Second, while there was certainly an overlap between inspection ratings and performance measures, this does not mean that data can simply replace inspections, and it is worth urging caution in this regard. While there might be relatively few 'false negatives' (poorly performing local authorities whose statistics indicate good performance), the history of child

protection shows that cases of unpredicted failure – or those presented by the media as such – can have a disproportionate impact (Jones, 2014; Ayre, 2001). Furthermore, inspection reports may contain contextual detail and an explanation of why problems have been experienced, information that statistics alone cannot provide. Third, while being found ‘inadequate’ by Ofsted certainly creates an impetus for change in affected local authorities – generally evidenced by an influx of extra funding – the findings here suggest that the main effect is an increased emphasis on CP interventions: more assessments, investigations, conferences and plans. For individual agencies censured by the inspectors, this could be an understandable strategy to re-assert managerial grip on the system, and may additionally reflect a reduction in early intervention and triaging. More troubling are signs in the analysis of trends that a similarly risk-averse mindset has permeated the sector as a whole (Hood et al., 2016; Devine and Parker, 2015). This mindset may have been precipitated by child abuse scandals but now seems to be locked into a regime of performance management driven by a combination of data and inspection.

Conclusion

Particularly in England, the Munro Review of child protection inspired hope that prescriptive and procedural controls over practice might be eased and renewed emphasis placed on expert analysis and skilled intervention (Munro, 2011). Subsequent years have seen a literal reduction in the volume of statutory guidance (HM Government, 2015) and a somewhat half-hearted shift towards more flexible timeframes for completing assessments (Munro and Stone, 2014). The findings discussed in this paper suggest that the issues highlighted by Munro are embedded not only in the available indicators but in approaches to performance management in the sector. The paucity of outcomes measures but also of evidence-based process measures have reinforced a dependence on procedural compliance backed up by a centralised inspectorate. Even without explicit targets, command and control regimes tend to subordinate a service’s purpose and design to its indicators, when the reverse should be true (Seddon, 2008). It is important to recognise that indicators simply indicate areas for deeper investigation, they do not prove a problem exists. The challenge for children’s services is to develop ‘double-loop learning’ (Munro, 2010: 15), i.e. to monitor and evaluate performance in a way that encourages innovation and better designed services. It is in helping local

authorities to carry out this function, rather than in punishing and rewarding their performance, that inspectors may find a more constructive role.

ACCEPTED MANUSCRIPT

References

Association of Directors of Children's Services (ADCS) (2015) *Comment on the Latest Ofsted Report into Children's Social Care*. Accessed online, <http://adcs.org.uk/inspection>

Ayre, P. (2001). Child protection and the media: Lessons from the last three decades. *British Journal of Social Work*, 31(6), 887-901.

Ayre, P. & Preston-Shoot, M. (eds.) (2010) *Children's Services at the Crossroads: A Critical Evaluation of Contemporary Policy for Practice*. Lyme Regis, Russell House Publishing

Baars, I. J., Evers, S. M., Arntz, A., & van Merode, G. G. (2010). Performance measurement in mental health care: present situation and future possibilities. *The International journal of health planning and management*, 25(3), 198-214.

Barth, R. P., & Jonson-Reid, M. (2000). Outcomes after child welfare services: Implications for the design of performance measures. *Children and Youth Services Review*, 22(9), 763-787.

Bevan, G., & Hood, C. (2006). What's measured is what matters: targets and gaming in the English public health care system. *Public administration*, 84(3), 517-538.

Cafcass (2012) *Three weeks in November... three years on... Cafcass care application study*. London: Cafcass

Cochrane, A. (2000). 'Local government: managerialism and modernisation'. In Clarke, J., Gewirtz, S. & McLaughlin, E. (eds.), *New Managerialism, New Welfare?*, London, Sage.

Department for Education (DfE) (2015a) *SFR 41/2015: Characteristics of children in need: 2014 to 2015*, accessed online, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469737/SFR41-2015_Text.pdf

Department for Education (DfE) (2015b) Statistical First Release, Children's Social Work

Workforce during year ending 2014, accessed online,

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/406660/Childrens_Social_Work_Workforce_2013-14

Devine, L., & Parker, S. (2015) *Rethinking child protection strategy: Learning from trends*, Bristol, Centre for Legal Research, University of the West of England.

Friedman, M. (1997). *A Guide to Developing and Using Performance Measures in Results-Based Budgeting*. Washinton, DC, Finance Project

Friedman, M. (2001). *The results and performance accountability implementation guide*. Available online, www.raguide.org

Gibson, J., & O'Donovan, B. (2014). The Vanguard Method as applied to the design and management of English and Welsh Children's Services departments. *Systemic Practice and Action Research*, 27(1), 39-55.

HM Government (2015) *Working Together to Safeguard Children: A guide to inter-agency working to safeguard and promote the welfare of children*. London: The Stationery Office

HM Government (2013) *Seizing the data opportunity: A strategy for UK data capability*. London: The Stationery Office

Hood, C. (1991). A public management for all seasons? *Public administration*, 69(1), 3-19.

Jones, R. (2014) *The story of Baby P: Setting the record straight*, London, Policy Press

Jones, R. (2015). The end game: The marketisation and privatisation of children's social work and child protection. *Critical Social Policy*, 35(4), 447-469

Lees, A., Meyer, E., & Rafferty, J. (2013). From Menzies Lyth to Munro: the problem of managerialism. *British Journal of Social Work*, 43(3), 542-558.

Lonne, B., Parton, N., Thomson, J., & Harries, M. (2008) *Reforming child protection*, Abingdon, Routledge.

Munro, E. (2004). The impact of audit on social work practice. *British Journal of Social Work*, 34(8), 1075-1095.

Munro, E. (2010). *The Munro review of child protection: Part 1: A systems analysis*. London: Department for Education

Munro, E. (2011). *The Munro review of child protection: Final report-a child-centred approach*. London: Department for Education

Munro, E. & Stone, J. (2014) The impact of more flexible assessment practices in response to the Munro Review of Child Protection. Accessed online, <https://www.gov.uk/government/publications>

Ofsted (2015a) Framework and evaluation schedule for the inspections of services for children in need of help and protection, children looked after and care leavers. Available online, <https://www.gov.uk/government/publications/inspecting-local-authority-childrens-services-framework>

Ofsted (2015b) *Annual report into children's social care 2013_14*. Accessed online, www.gov.uk/government/organisations/ofsted

Osbourne S., Bovaird, T., Martin, S., Tricker, M., Waterson, P. (1995) Performance Management and Accountability in Complex Public Programmes. *Financial Accountability and Management*, 11: 19-37

Propper, C., & Wilson, D. (2003). The use and usefulness of performance measures in the public sector. *Oxford review of economic policy*, 19(2), 250-267.

Rothstein, H., Huber, M. & Gaskell, G. (2006). 'A theory of risk colonization: The spiralling regulatory logics of societal and institutional risk'. *Economy and Society*, 35(1), 91-112.

Sass, J., & Crosbie, T. (2013). Democracy and Scandal: A Research Agenda. *Comparative Sociology*, 12(6), 851-862.

Seddon, J. (2008). *Systems thinking in the public sector*. London: Triarchy Press.

Shaw, I., Bell, M., Sinclair, I., Sloper, P., Mitchell, W., Dyson, P., & Rafferty, J. (2009). An exemplary scheme? An evaluation of the Integrated Children's System. *British Journal of Social Work*, 39(4), 613-626.

Smith, P., Mossialos, E., & Papanicolas, I. (Eds) (2009). *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press.

Tilbury, C. (2004). The influence of performance measurement on child welfare policy and practice. *British Journal of Social Work*, 34(2), 225-241.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301-320.

Highlights

- Quality measures mainly focused on the timeliness of work processes, but these did not seem to affect outcomes in the form of re-referral rates.
- Re-referrals, which are a negative outcomes measure, were higher in local authorities with a tendency to close cases quickly and in those with high rates of agency workers.
- A small number of indicators were often enough to predict an inadequate Ofsted rating in 2012 and 2013. These indicators were: timeliness of initial assessments, rates of re-referrals, and agency worker rates
- Changes in performance measures in the year following an inadequate Ofsted rating seemed to suggest greater use of protective interventions compared with similarly performing local authorities.
- The Ofsted inspection process, with high numbers of local authorities rated inadequate, may be exacerbating the shift towards protective interventions in the sector.