Measurement and improvement of emergency department performance through inspection and rating: an observational study of emergency departments in acute hospitals in England

Thomas Allen, Kieran Walshe, Nathan Proudlove, Matt Sutton

Department of Public Health and Policy, University of Manchester, Manchester, UK

ABSTRACT

Introduction Hospital inspection and the publication of inspection ratings are widely used regulatory interventions that may improve hospital performance by providing feedback, creating incentives to change and promoting choice. However, evidence that these interventions assess performance accurately and lead to improved performance is scarce.

Methods We calculated six standard indicators of emergency department (ED) performance for 118 hospitals in England whose EDs were inspected by the Care Quality Commission, the national regulator in England, between 2013 and 2016. We linked these to inspection dates and subsequent rating scores. We used multilevel linear regression models to estimate the relationship between prior performance and subsequent rating score and the relationship between rating score and post-inspection performance.

Results We found no relationship between performance on any of the six indicators prior to inspection and the subsequent rating score. There was no change in performance on any of the six indicators following inspection for any rating score. In each model, CIs were wide indicating no statistically significant relationships.

Discussion We found no association between established performance indicators and rating scores. This might be because the inspection and rating process adds little to the external performance management that EDs receive. It could also indicate the limited ability of hospitals to improve ED performance because of extrinsic factors that are beyond their control.

INTRODUCTION

Many countries use hospital inspections and published ratings to measure and improve performance, but little is known about the likely impact. Such systems of regulation, including inspection, accreditation and certification, have been the topic of several reviews. Brubakk et al. found no evidence in support of the effectiveness of accreditation and concluded that quality of care was not improved by this approach. Hinchcliff et al. noted the lack of robust studies in this area and concluded that further research is needed. Flodgren et al. found only two studies and therefore drew no conclusions on the impact of external inspections on quality standards.

In England, hospital inspection and rating are conducted by the Care Quality Commission (CQC). They inspect all hospitals in England and undertake focused inspections in emergency departments (EDs). They rate the performance of EDs on a four-point scale (outstanding, good, requires improvement or inadequate) in five domains (effectiveness, safety, care, responsiveness and leadership) and also assigned an overall rating.

In England, the performance of EDs is also measured through a set of performance indicators that are defined and used by the National Health Service (NHS) to provide information on the quality of care of EDs to help support quality improvement. Performance on these indicators is published every month (box 1). All indicators relate to first and unplanned follow-up attendances. Indicators one to five decrease with better ED performance. Indicator six increases with better performance.

CQC inspections of EDs follow a detailed inspection handbook, which sets out the areas of investigation to be explored, what information should be gathered in assessing performance and the process for assigning ratings. The inspection handbook covers some of the same issues as the NHS England performance indicators, including waiting times for...
ED performance indicators.

1. Time to initial assessment: The median time (in minutes) between the patient’s arrival and their initial assessment.
2. Time to treatment: The median time (in minutes) between the patient’s arrival and the start of their treatment.
3. Total time spent in the emergency department: The median time (in minutes) between the patient’s arrival and departure from the emergency department.
4. Left department before being seen for treatment: the percentage of attendances with a ‘disposal-type’ of Left Before Treatment.
5. Unplanned re-attendance: the percentage of attendances for which there was an unplanned re-attendance within 7 days.
6. Total time spent in the emergency department ≤4 hours: the percentage of patients for whose total time spent in the emergency department is less than 4 hours.

ED, emergency department.

Box 1

The selected examples in table 1 show, the inspection reports may refer directly to the indicator metrics. These report extracts suggest that we may expect to see a relationship between the CQC ratings and the NHS England performance indicators, but the reports have not been identified systematically. In this paper, we examined whether those EDs which do well on the NHS England performance indicators also perform well when inspected and rated by CQC, and vice versa. Further, we examined whether, after a CQC inspection, performance on the NHS England indicators improves, especially for EDs which were rated inadequate or requires improvement, as these departments probably have both the greatest scope for improvement and the strongest incentive to do so. To test these hypotheses, we examined whether prior levels of performance on six indicators were associated with the ratings they received when inspected. We also analysed whether levels of performance on those indicators changed following inspection and rating, for all EDs and specifically for providers who were rated as requires improvement or inadequate.

METHODS

Data

We used inspection data from the CQC9 and data on hospital activity10 to analyse the associations of inspections and ratings between 2013 and 2016 with hospital performance in England.

CQC provided information on the inspection date, the rating scores and when the inspection report was published for all first inspections of ED services in England (207 in total). We used the results from first inspections as these are large-scale and inspect more aspects of care than subsequent re-inspections, which are smaller and more targeted.

We used activity data from EDs between April 2012 and September 2016 to generate results for the six indicators that are routinely used to measure ED performance. The indicators were created for type 1 EDs only. These are departments that provide 24-hour consultant-led care. Single-specialty and minor injury units are excluded.

Indicators measuring time to initial assessment, time to treatment and total time in the ED were cleaned to remove values where the time was either zero or 24 hours. These values are highlighted by NHS Digital in the indicator methodology documentation as suggesting issues with data quality.8 From 65 million assessment, treatment and admission, and unplanned re-attendance. As the selected examples in table 1 show, the inspection reports may refer directly to the indicator metrics.

Table 1

| Hospital                        | Extract from report narrative                                                                 |
|--------------------------------|-----------------------------------------------------------------------------------------------|
| Trust rating: outstanding      | ► The department exceeded the target of 95% of all patients to be admitted, transferred or discharged within 4 hours of arrival to the ED every month. The trust had been meeting this target annually since February 2012 and was one of the top five performing trusts in the country. ► The proportion of patients leaving before being seen was lower than the England average. ► The unplanned re-attendance rate to the ED within 7 days of discharge ranged between 5% and 7%. This was consistently lower than the England average of above 7% but higher than the national standard of 5% |
| Trust rating: good             | ► Since July 2015, the department had met the national 4-hour waiting time target and most patients were discharged within 3 hours of attendance. The trust was performing better than the England average for a number of other performance measures relating to the flow of patients. ► Between June 2015 and September 2015, the trust rate for unplanned re-attendance at A&E within 7 days was 0.3% (better than the England average). ► Only 3.2% of patients left the department before a clinician saw them. This was significantly better than the 5% standard set by the trust. |
| Trust rating: requires improve | ► The trust performed mostly above the 95% standard for percentage of patients waiting to be seen within 4 hours since May 2014, with the exception of December 2014 and May 2015. This was an improvement in the previous years. ► The latest available information showed that the unplanned re-attendance rate for September 2014 was 2.2%, while in August it was 2.5%, and 2.7% in July. ► The number of patients leaving without being seen was higher than the national average. In March 2015, this was 2%, while in February, it was 1.8%, and in January, 2%. This was the latest available information. |
| Trust rating: inadequate        | ► National standards of being admitted, transferred or discharged within 4 hours had not been met since October 2014. The processes put in place to trigger action to deal with poor flow through the ED were delayed and slow, and patients frequently and consistently could not access the hospital in a timely way, experiencing unacceptable waits. ► The rates of unplanned re-attendance for the trust were lower than the national average for most of 2013 to 2015 but were higher for February 2015. We asked what had been identified as the reason for the increase but the management of the ED could not offer any specific reason and the increase had not been investigated. ► The percentage of patients leaving before being seen was higher than the national average for most of January 2013 to October 2015. |

CQC, Care Quality Commission; NHS, National Health Service.
ED attendances, the cleaning resulted in 6.67% of attendances removed for a time to initial assessment, 1.48% removed for a time to treatment and 0.06% removed for total time in the ED.

CQC inspected and rated ED services at 207 individual hospital ED sites. The hospital activity data, and therefore our indicators, were available for 150 hospital trusts, some of which include multiple hospital sites. Our analysis could only be conducted where the two data sources could be linked, that is to say, for single site trusts or multi-site trusts with the same rating for all their sites. We linked the data at hospital trust level keeping 118 trusts consisting of 86 single site trusts and 32 multisite trusts with the same inspection date and rating for all their sites.

**Analysis**

We examined whether levels of performance before the inspection cycle started differed according to the subsequent rating score. This shows if EDs that are judged to be better by the CQC also tend to have shorter waiting times, fewer walk-outs before treatment and fewer re-attendances. We only considered performance in the period April 2012 to September 2013, before the inspection cycle started.

We also examined whether the indicators of ED performance changed after the services were inspected. We allowed for differential effects depending on the rating score and examined whether there were changes shortly before the inspection and in the short term and long term after the inspection. We used a multilevel linear regression model with indicators for trust and month. Observations were weighted by the volume of emergency attendances. We modelled the performance indicators using two sets of variables indicating the period following the inspection and the rating that the service received. The short-term post-inspection period covers the inspection month and the 6 months immediately following. This period differs across hospitals as inspections occurred at different times.

Additional models allowed trusts to respond 1 month prior to inspection (an anticipation effect) and to respond >6 months after the inspection (longer-term effect). These analyses included data until September 2016 and therefore, the follow-up period was dependent on the timing of the inspection.

All models were estimated on trust-level data. We used separate multilevel models for each performance indicator with trust effects, to account for unobserved factors affecting performance, and indicators for a month, to account for seasonal effects. The volume of emergency attendances per trust per month was dependent on the timing of the inspection period. Rating scores were evenly distributed between these two rating outcomes and over the inspection period.

**RESULTS**

In the period before the inspection cycle started, on average, the hospitals performed an initial assessment after 6 min, treated patients after 59 min and kept patients in the ED for a total of 145 min (Table 2). 2.81% of attendances ended with the patient leaving the department before being seen and 5.91% of patients returned to the ED within 7 days. The average percentage of patients achieving the target of <4 hours spent in the ED was 94%.

Figure 1 displays the number of ED ratings, grouped as either positive or negative, and their corresponding inspection month. Rating scores were evenly distributed between these two rating outcomes and over the inspection period.

Figure 2 presents the mean values of each indicator from April 2012 to September 2013 split by subsequent rating score. Across these six indicators, there are some isolated cases where performance appears to differ by subsequent rating score. However, overall, there is no clear pattern in these differences. This finding is confirmed in Table 3 when indicator performance was regressed on the rating score. In Table 3, the coefficients represent the difference in indicator performance, measured in minutes or percentage points, observed prior to inspection for those departments with a positive rating compared with those with a negative rating. For example, the results indicating EDs with a positive rating had shorter times to assessment (1.71 min) in the pre-inspection period. Similarly, EDs with a positive rating had 0.15% more patients leaving without being seen. However, across all six indicators, the difference is only statistically significant for performance on the 4-hour target and suggests that EDs with positive ratings performed better on this indicator.

Table 4 presents the results from our main model specification focusing on the post-inspection period. For each of the ED performance indicators, coefficients are presented in either minutes for time-dependent indicators or percentage for all others. There were no statistically significant changes in these coefficients and hence no detected change in ED performance after services had been inspected, regardless of the inspection rating they received.

| Table 2 Summary statistics for ED indicators |
|---------------------------------------------|
| Median | IQR  | SD     | Minimum | Maximum |
|---|---|---|---|---|
| Median time to initial assessment (minutes) | 6 | 5 | 17.97 | 1 | 123 |
| Median time to treatment (minutes) | 59 | 19 | 58.38 | 9 | 571.50 |
| Median time spent in ED (minutes) | 145 | 33 | 22.90 | 83 | 199 |
| Left department before being seen for treatment (percentage) | 2.81 | 1.37 | 1.15 | 0.40 | 6.67 |
| Unplanned re-attendance (percentage) | 5.91 | 1.65 | 1.27 | 2.97 | 9.03 |
| Total time spent in ED <4 hours (percentage) | 93.81 | 3.70 | 4.45 | 71.35 | 98.62 |

Unweighted averages of trust-level data for 118 trusts. Over the pre-inspection period: April 2012 to September 2013.
In table 5, we present the results including terms capturing an anticipation effect and a longer-term post-inspection effect (>6 months). There is weak evidence of anticipation effects, indicating that services that subsequently had inadequate and requires improvement ratings deteriorated on the 4-hour target in the month prior to inspection. There was no statistically significant change over the longer time period.

Similarly, three different supplementary analyses found no significant relationship between indicators and inspection ratings. These analyses included outlier values for waiting times, the original four category rating score and the publication date of the rating report.

DISCUSSION
Summary of findings
In this study, we failed to detect a relationship between indicator performance and CQC inspection ratings during the period 2013–2016 in England, either before inspection or post-inspection. This does not exclude the possibility that a relationship exists, but that we were unable to detect this in our analysis.

In the period before CQC inspections began, there was considerable variation in the performance of EDs on the six indicators analysed in this study. However, this variation was not associated with subsequent rating scores; that is, we did not find evidence that services with different inspection rating scores systematically performed differently before inspection. Because the CQC inspection process gathers a much wider and richer dataset than the six indicators, we would not necessarily expect a strong relationship between ED indicator performance and subsequent scores. However, we would have expected to see some association. The absence of such an association raises questions about the validity or reliability of the ED performance indicators or the CQC inspection ratings (or both) as measures of quality.

Using the NHS England indicators, we also found no evidence that, after inspection, EDs changed their performance. This is both when looking at the short-term (6 months) and longer-term (>6 months) post-inspection effects. It may be that the performance management regime in EDs, with intense attention from both NHS England and NHS Improvement, combined with the publication of regular performance statistics, leaves little room for CQC’s inspection process to generate additional improvement. It might also suggest that the variations in performance result from extrinsic factors (such as the availability of primary care services, alternatives to EDs, such as walk-in clinics and social care provision) rather than from intrinsic factors more under the control of hospitals.

Strengths and weaknesses
Our study focuses on the first round of a new inspection regime in England and therefore contributes to our understanding of the impact of large-scale and targeted inspections. We chose to look at ED services because they are a distinctive and separate clinical service with a substantial amount of available data regarding performance and because they are separately inspected, rated and reported on by CQC. In doing so, we ensure that our measures of performance are as targeted as they can be. In the absence of a randomised control trial, we cannot rule out all forms of bias, but our statistical methods have been selected to minimise its impact. Our analysis takes advantage of the timings of inspection that are not random but plausibly exogenous to service performance, and we controlled for seasonal and macro effects. We further controlled for ED demand and weighted our models by patient volumes.

Our study is limited as we cannot link activity data and ratings for all EDs. We had to exclude 46 hospitals across 20 hospital trusts. The distribution of rating scores is broadly similar between the departments included in and excluded from the study.
Table 3  Differences in indicators in the preinspection period

|                      | Median time to initial assessment (minutes) | Median time to treatment (minutes) | Median total time spent in ED (minutes) | Left department before being seen for treatment (percentage) | Unplanned re-attendance (percentage) | Total time spent in ED < 4 hours (percentage) |
|----------------------|--------------------------------------------|-----------------------------------|----------------------------------------|-------------------------------------------------------------|--------------------------------------|---------------------------------------------|
| Positive rating      | $-1.713 \pm (-12.39 \text{ to } 8.96)$    | $-2.854 \pm (-16.24 \text{ to } 10.53)$ | $-1.077 \pm (-7.89 \text{ to } 5.74)$ | $0.159 \pm (-0.17 \text{ to } 0.49)$                        | $0.121 \pm (-0.27 \text{ to } 0.51)$   | $1.610^* \pm (0.26 \text{ to } 2.96)$         |
| Observations         | 1984                                       | 2054                              | 2108                                   | 2108                                                        | 2108                                 | 2108                                        |
| Trusts               | 114                                        | 115                               | 118                                    | 118                                                         | 118                                  | 118                                         |
| $R^2$ (within)       | 0.009                                      | 0.023                             | 0.189                                  | 0.126                                                      | 0.155                                | 0.195                                       |

Reference category: negative rating (inadequate or requires improvement). Random effects regression. Robust 95% CI are within parentheses. Models also include 18 indicators for each month of the preinspection period (April 2012 to September 2013) and the number of emergency attendances. Observations are weighted by number of emergency attendances. *p<0.05, **p<0.01, ***p<0.001.

ED, emergency department.

Figure 2  ED indicators for each rating score. ED, emergency department.
Table 4 Changes in ED indicators post-inspection by rating score

| Six months post-inspection effect for | Median time to initial assessment (minutes) | Median time to treatment (minutes) | Median total time spent in ED (minutes) | Left department before being seen for treatment (percentage) | Unplanned re-attendance (percentage) | Total time spent in ED <4 hours (percentage) |
|--------------------------------------|-------------------------------------------|-----------------------------------|----------------------------------------|----------------------------------------------------------|-------------------------------------|-----------------------------------------|
| Negative rating                      | −6.972 (−20.42 to 6.47)                   | −0.185 (−5.85 to 5.48)            | −0.198 (−3.75 to 2.68)                | −0.198 (−0.22 to 0.06)                                    | −0.0347 (−0.29 to 0.28)            | −1.117 (−2.33 to 0.10)                 |
| Positive rating                      | 14.65 (−9.54 to 38.84)                   | 1.329 (−4.36 to 7.02)             | 2.963 (−5.64 to 11.56)               | 0.0865 (−0.27 to 0.10)                                    | −0.00543 (−0.29 to 0.28)           | −0.693 (−2.70 to 1.31)                |

Observations

| Trusts                          | 4486 | 4662 | 4758 | 4748 | 4758 | 4758 |
|---------------------------------|------|------|------|------|------|------|
| nested trust                    | 0.517| 0.442| 0.643| 0.691| 0.518| 0.581|

Table 5 Changes in ED indicators preinspection and post-inspection by rating score

| One month pre-inspection effect for | Median time to initial assessment (minutes) | Median time to treatment (minutes) | Median total time spent in ED (minutes) | Left department before being seen for treatment (percentage) | Unplanned re-attendance (percentage) | Total time spent in ED <4 hours (percentage) |
|-------------------------------------|-------------------------------------------|-----------------------------------|----------------------------------------|----------------------------------------------------------|-------------------------------------|-----------------------------------------|
| Negative rating                     | −1.925 (−10.42 to 6.57)                  | −3.26 to 6.85                    | 3.090                                  | −0.22 to 6.40                                            | −0.18 to 0.29                       | −0.00321                               |
| Positive rating                     | 9.090 (−25.75 to 45.56)                  | 1.706                            | 8.678                                  | −9.66 to 27.01                                           | −0.21 to 0.22                       | −0.026                                 |

Six months pre-inspection effect for

| Negative rating                     | −6.272 (−22.08 to 9.54)                  | 0.158                            | 0.495                                  | −2.86 to 3.84                                           | −0.42 to 0.07                       | −0.0402                               |
| Positive rating                     | 11.64 (−9.53 to 32.81)                   | 2.063                            | 4.065                                  | −5.07 to 13.20                                          | −0.26 to 0.15                       | −0.288                                |

Remaining months post-inspection effect for

| Negative rating                     | 1.210 (−18.84 to 21.26)                  | 1.182                            | −1.224                                 | −6.85 to 4.40                                          | −0.52 to 0.29                       | 1.103                                 |
| Positive rating                     | 14.62 (−21.07 to 50.31)                  | 5.951                            | 0.0737                                 | −4.46 to 4.61                                          | −0.27 to 0.30                       | 0.00911                               |

| Observations                        | 5929 | 6155 | 5267 | 5267 | 5267 | 5267 |
|-------------------------------------|------|------|------|------|------|------|
| nested trust                        | 0.485| 0.494| 0.669| 0.624| 0.508| 0.623|

Policy implications

Our study suggests that the effects of CQC’s inspections and ratings are not measurable using the metrics widely used by NHS England and the Department of Health and Social Care in performance management. This could result from four possible scenarios: (1) the chosen indicators are valid measures of quality, but CQC inspection ratings are not; (2) the chosen indicators are not valid measures of quality, but the CQC inspection ratings are; (3) both are valid measures of quality but they measure different aspects or dimensions of quality; or (4) neither are valid measures of quality.

CQC intends to move towards a more intelligence-led approach to regulation,11 and such an approach will rely more heavily on routine data such as the indicators used in this study. Our findings suggest that CQC may wish to consider developing new measures of performance or using existing measures differently.

Our findings suggest that when future regulatory interventions such as inspection programmes are being developed, they should be piloted and tested to evaluate their association with other performance measures and to assess their effects before being widely implemented.

Correction notice Since this article was published online, the text from figure 1 has been added to the text and figure 2 and 3 have been renumbered as figure 1 and 2.

Contributors TA, KW, NP and MS: design of this study. TA: conducted the analysis. TA, KW, NP and MS: drafted this article.

Funding This research was funded by the NIHR Policy Research Programme (PR-R11-0914-12001). The open access funding was provided by the Institute for Health Policy and Organisation, University of Manchester.

Disclaimer The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Competing interests None declared.
REFERENCES
1 Greenfield D, Braithwaite J. Developing the evidence base for accreditation of healthcare organisations: a call for transparency and innovation. Qual Saf Health Care 2009;18:162–3.
2 Hinchcliff R, Greenfield D, Moldovan M, et al. Narrative synthesis of health service accreditation literature. BMJ Qual Saf 2012;21:979–91.
3 Flodgren G, Pomey MP, Taber SA, et al. Effectiveness of external inspection of compliance with standards in improving healthcare organisation behaviour, healthcare professional behaviour or patient outcomes. Cochrane Database Syst Rev 2011:CD008992.
4 Brubakk K, Vist GE, Bukholm G, et al. A systematic review of hospital accreditation: the challenges of measuring complex intervention effects. BMC Health Serv Res 2015;15:280.
5 NHS Digital. Provisional Accident and Emergency Quality Indicators - England. 2017 http://digital.nhs.uk/catalogue/PUB23839.
6 NHS Digital. Methodology for producing the A&E clinical quality indicator from provisional Hospital Episode Statistics (HES) data. 2016 https://files.digital.nhs.uk/publicationimport/pub23xxx/pub23839/prov-a-e-qual-indi-eng-january (accessed 31 Jan 2018).
7 NHS England. A&E Attendances and Emergency Admissions Monthly Return Definitions. 2015 https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/03/AE-Attendances-Emergency-Definitions-v2.0-Final.pdf (accessed 31 Jan 2018).
8 Care Quality Commission. Inspection framework: NHS and independent acute hospitals. Core service: Urgent and emergency services. 2018 http://www.cqc.org.uk/sites/default/files/inspection-framework-acute-hospitals-urgent-emergency-services.pdf (accessed 7 Jun 2018).
9 Care Quality Commission. How to get and re-use CQC information and data. 2016 http://www.cqc.org.uk/content/how-get-and-re-use-cqc-information-and-data (accessed 21 Jun 2016).
10 NHS Digital. Hospital Episode Statistics. 2018 http://content.digital.nhs.uk/hes (accessed 31 Jan 2018).
11 Care Quality Commission. Shaping the future. CQC’s strategy for 2016 to 2021. 2016 http://www.cqc.org.uk/sites/default/files/20160523_strategy_16-21_strategy_final_web_01.pdf (accessed 1 Sep 2016).