“Gaming the System”: Testing the Validity and Reliability of NHS Inspections of Hospital Cleanliness

Veronica Toffolutti, postdoctoral researcher[1], Martin McKee, professor of European public health[2], David Stuckler, professor of political economy and sociology[1,2]

1. Department of Sociology, Oxford University, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK
2. Department of Public Health and Policy, 103 London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH UK

Correspondence to: V. Toffolutti
veronica.toffolutti@sociology.ox.ac.uk

Keywords: Hospital Inspections; Hospital Cleanliness; Gaming; Opportunistic behaviour

Word Count: 3,697

Abstract

Audits are a key means to monitor and ensure quality of care and maintain high standards in the English NHS. Yet there is a perception that they can be gamed. This can happen, for example, when staff know that an audit will soon take place.

Using a Regression Discontinuity Design on data for 205 English NHS hospitals, covering the period 2011-2014, we tested whether perceptions of cleanliness increased during periods when inspections occurred.

Our results show that during the period within 2 months of when hospitals were being inspected we observed a significant elevation in patients reporting greater cleanliness, by around 9%. This association was consistent even after adjusting for secular time trends. These associations were concentrated in hospitals which outsource cleaning services and were not detected in those using NHS cleaning services.

Introduction

A pre-requisite for a competitive market in health care, such as that established by the English 2012 Health and Social Care Act, is the existence of valid information on the performance of providers. This is necessary for informed purchasing of services. Yet, as has long been noted, this can be difficult because, other than for certain easily standardised services, many aspects of health care are difficult to specify(1) and there are strong incentives for opportunistic behaviour, or gaming.(2) This can take many forms, such as changing behaviour, for example by avoiding complex cases, or changing how things are recorded, such as adding diagnostic codes to make patients appear more severely ill than they are.(3)

One area of concern relates to external inspections of providers, such as those undertaken by one of the regulators in the NHS in England, the Care Quality Commission, as noted on several occasions by the House of Commons Public Accounts Committee.(4) These concerns are echoed in education, which has also seen a marked increase in inspections and where there have been many accounts of opportunistic behaviour,(5) such as schools being warned about “unplanned” inspections or the temporary exclusion of disruptive or low ability students from testing(6-8) or even changing the food provided in school cafeterias with the dubious intention of boosting students’

performance(9) (with questionable impacts on their health).(10, 11)

Hospital cleanliness has been high on the agenda of successive governments in the United Kingdom, reflecting a combination of both appropriate concern about hospital acquired infection and the exploitation of data by some media outlets.(12) Even though the intensity of coverage has diminished, it has not stopped. (13-17)

Consequently, the NHS Ten-Year Plan, launched in 2000,(18) established a series of “nation-wide clean-up campaigns” to improve cleanliness in hospitals. These involved ‘unannounced’ inspections (although staff were always given 48 hours’ notice) which would take place over the course of up to one month, by teams composed initially of hospital staff and patients. However, a lack of patient volunteers meant that they subsequently drew mainly on NHS staff.

From the outset, there has been concern about the potential for gaming of cleanliness audits. It is widely believed that, since staff know when each inspection will happen, they are incentivised to make a special effort in the period leading up to it, subsequently relaxing their standards. It might be that this would be especially prominent in services which are outsourced to private contractors, given the risk of failing to achieve contract renewal should their performance receive poor scores in NHS inspections.

The true extent and consequences of gaming in the NHS are poorly understood but there is enough evidence to raise concerns, with Mannion and Braithwaite finding 20 distinct forms of dysfunctional responses to the NHS performance management regime. (19) Bevan and Hood give examples of poor performance in areas not measured, hitting the target but missing the point, and ambiguities or fabrication of data.(20) Another review of responses to targets identified creation of target-free zones, either physically (e.g. awaiting admission in temporary facilities in hospital car parks) or administratively (e.g. informal “waiting lists” to get on official waiting lists), and exploiting the opportunity to remove patients from waiting lists if they decline an offer of admission by making offers during holiday periods.(21)In addition, two studies found that financial incentives to physicians increase the likelihood that they will manipulate lists of patients by excluding those whose presence impedes their achievement of targets. (22, 23)

In these circumstances, it seems plausible that gaming of audits of hospital cleaning, cannot be excluded; information we obtained from two acute Trusts under Freedom of Information legislation revealed that they actually had between two and five months’ notice of inspections.

Here, for the first time to our knowledge, we look for evidence of possible ‘gaming effects’, by taking advantage of

a unique source of data which links patients' perception of cleanliness, with hospital audits date covering the years 2011-2014. Specifically, we test whether patients report higher cleanliness in the months leading up to an inspection than at other times, which would be consistent with the hypothesis that gaming does take place.

Data and Methods

We linked data on patients' perceptions of cleanliness with dates of hospital cleaning inspections for 205 English hospitals. All analyses are conducted at hospital level. Patient-reported cleanliness data were obtained from the Picker Institute NHS Patient Survey Programme.(24) Each Trust (public organisation operating one or more health care providers, including hospitals in England) sends a questionnaire to 850 patients who have spent at least one night in hospital between June and August each year. They are asked to report on experiences at any time in the year although, in practice, 93% of reports describe experiences in this three-month period. All the sampled patients were asked "In your opinion, how clean was the hospital room or ward that you were in? Very clean (excellent), fairly clean, not very clean, not clean at all". We re-coded the data by hospital, and matched this with the month

of audit, obtained from Patient Environment Action Teams (PEAT)(2011-2)(25) and Patient-Led Assessments of the Care Environment (PLACE)(2013-4) (note: the name changed but collection practices did not).(26) We aggregated these to measure the median percentage of patients giving an excellent cleanliness score for each hospital by month and year.

Additional data on size and services provided by the hospitals were taken from the Estates Return Information Collection (ERIC) for the period 2011-2014.(27)

We matched data based on the calendar year for PEAT/PLACE and NHS Inpatient Survey. ERIC data on hospital beds, however, report in financial year terms, which we matched to calendar years. This is unlikely to confound the analysis since there is little temporal variation in numbers of hospital beds.

Our initial sample included 492 English hospitals. 17 (3.46%) were excluded because they had no inpatient services. Another 270 (54.9%) were excluded because patients had not been surveyed. Thus, the final sample consists of 205 English hospitals and a total of 907 hospital-months. Of these 125 hospitals operated in-house, NHS cleaning services, and 76 hospitals contracted with private providers, 4 operated with both NHS and private providers (i.e. hospitals which integrated outsourcing into a mixed public-private partnership). Web Appendix 1 Exhibit A1 (28) displays this in flow chart form.

Exhibit 1 provides further descriptive statistics for all variables used in the study.

[Exhibit 1 about here]

Statistical Models

To investigate the association between month of inspection and perceived cleanliness we use a Regression Discontinuity Design, (29) further details presented in Web Appendix 1 Exhibit A2.(28)

As shown in Web Appendix 1 Exhibit A3, (28) until 2012 (PEAT survey) the assessments tended to concentrate between January and March, whereas the PLACE assessments tended to span the first six months of the year. The main coefficient of interest is β, which estimates the average change in the median perceived cleanliness for hospitals during inspection months. All data and models were estimated using Stata version 13. Robust standard errors were clustered by hospital to reflect non-independence of sampling.

Limitations

As with all statistical modelling studies, our analysis has several limitations. First, we do not have the exact date when the patient was discharged but only the month so, when we merge information at site level, we cannot investigate a possible gaming effect within the month. This imprecision is likely to

produce conservative estimates of the magnitude of potential gaming behaviour. Second, although the magnitude of effect size may appear modest, this increment in perceived cleanliness is sufficient for sites to avoid threats of an adverse assessment and the consequences that flow from it. Third, ideally a comprehensive longitudinal dataset would be available that tracks independent patient perceptions of cleanliness across the entirety of the sites in the UK. This does not exist, so in this initial assessment, to our knowledge for the first time in the NHS, we have taken advantage of a large pooled dataset to observe if there is an elevation in cleanliness in the months just in advance of and during inspections, which reverts to its historic level after the inspection. A limitation is that this cannot identify individual hospital which are gaming but does point to characteristics, such as outsourcing, which may render them more likely to do so. Fourth, we cannot observe a uniform distribution both in terms of month of assessment nor in terms of patients responding to the questionnaire. This is a strength, but also a limitation, in that we are taking advantage of these available data to assess gaming effects.

Our data are corroborated by other evidence. The FOI requests to sites about communication with cleaning staff in the months when inspections were undertaken reveal that detailed pre-assessment checks, revealing longstanding problems that then

be addressed, were undertaken a few days before the inspections (see Web Appendix 1, Exhibit A4). (28)

Results

Association of Inspection Months with Cleanliness

Exhibit 2 presents the median trend in patients reporting excellent hospital cleanliness in the periods preceding, during, and following an NHS cleanliness inspection. In the months approaching inspection dates, cleanliness appears to rise, and then drops after the inspection period. Comparing the months before and after the inspection, on average, patient reports of excellent cleanliness are about 10 percentage points higher (81.5% in inspection months versus 71.9% in all other months, t-test comparison= -3.73 p-value <=0.001).

To provide an illustrative example, exhibits 3a and 3b depict the trend for the Royal National Hospital for Rheumatic Diseases. Inspections were in June 2013 and May 2014. In the months before each inspection, patient perceptions of cleanliness are relatively constant. Then, in the inspection month it jumps, and shortly after returns to the prior level.

[Exhibit 2 and 3a and 3b about here]

Exhibit 4 shows the results of our statistical models. In those months when an inspection took place the reported cleanliness rate jumped by 7.78 percentage points (95% CI: 2.75% to 12.8%). To further corroborate our results we present visually the estimation coefficients of distributed lag model in Web Appendix 1, Exhibit A5. (28)

[Exhibit 4 about here]

Comparing Hospitals with Outsourced Cleaning Services to NHS In-house Services

We applied a difference-in-differences model to test if those cleaning services operated by private actors were more likely to exhibit gaming behaviour.(30) As shown in Exhibit4 column 2, higher cleanliness scores in inspection months were concentrated in hospitals that outsource cleaning services (11.0percentage points, 95% CI: 5.15% to 19.6%), whereas there was no statistically detectable association in those hospitals using in-house NHS cleaning services (2.68 percentage points, 95% CI: -3.52% to 8.88%). To further corroborate our results we present visually the estimation coefficients of distributed lag model in Exhibit A6 in Web Appendix 1. (28) This is line with a recent research that finds a greater incidence of

infection and evidence of poorer cleaning where it is outsourced.(31)

Fixed Effect Estimation

To test whether our results were driven by potential unobserved heterogeneity we used a within-group estimation. Our results clearly show that switching from a non-assessment month to assessment one leads to an increase in the reported cleanliness by about 2.54 percentage points (95% CI: 0.02% to 5.06%).

To further corroborate our results, we included a cubic term in the term “time to the inspection”, finding consistent results (β=2.86%, 95% CI: 0.06% to 5.67%).

Robustness Checks

We performed a series of robustness checks. First, we adjusted for potential confounding factors, including hospital size, complexity and time-trends. All results are presented in exhibit A7 in Web Appendix 1. (28)

To identify whether these patterns were driven by a few outliers exhibiting extreme gaming activity, we removed 5% of our distribution (2.5% respectively from the bottom and the top of the distribution). As shown in exhibit A8 of Web Appendix 1, none of the results was changed. (28)

We further examined whether our results were confounded by some areas having low numbers of respondents. We restricted our sample to those areas with at least three hospitals with at least 17 respondents, so removing 10% of the lower end of the distribution in terms of number of respondent patients for each month (presented in Web Appendix 1 exhibit A9).(28) The results were consistent with our main findings, except for the fixed effect results, which are no longer significant.

To ensure that our results were not driven by difference between acute and specialist hospitals, we apply two different robustness tests, presented in Web Appendix 1 Exhibit A10. (28) First, we use propensity score matching (PSM) to better match hospitals observed during assessment months and afterwards. This reduces potential confounding by comparing hospitals with matching size and complexity. More precisely, we defined as treated those sites observed during assessments months or shortly before and controls as the others. We stratify by type of hospital (specialist, multi-service or acute), hence generating three cells. Within each cell, we align the distribution in terms of hospital size via PSM (1 to many, with replacement). As goodness of fit we imposed a 1% caliper, meaning that those matching couples which have a difference in their propensity score larger than 0.01 were automatically discarded. Second, we restrict our sample only to specialist sites. In both cases, none of our results changed qualitatively.

Finally, to further corroborate our gaming hypothesis we analysed the response pattern to food quality (Exhibit A11 in Web Appendix 1)(28). It is worth noting that the sample size drops because the assessment of food quality is available at Trust level only. Moreover, this is a good test on conceptual and empirical grounds. Conceptually, cleaning and food are different services. The companies providing these services are also different, if outsourced. This makes it a specific test that it is cleaning not a general disposition to outsourcing that is problematic. Empirically we observe no significant correlation at trust level of cleanliness and food and hydration quality scores (ρ=0.11).

Discussion

NHS inspections are a core element of the regime designed to ensure that hospitals maintain high standards of quality. This is especially important when services, including cleaning, are outsourced to private contractors to save money. Yet, there is a perception that NHS inspections can be gamed. This can happen, for example, when staff know that an inspection will soon take place.

By taking advantage of a unique data source, we can compare patient perceptions around the time of inspections. We find

evidence consistent with gaming: in audit months and in a short period before, cleanliness appears to rise, and then falls in subsequent months. This pattern was most prominent for hospitals which outsource cleaning services to private contractors. This appears particularly relevant since a recent study finds that sites with outsourced cleaning services have significantly higher rates of *methicillin-resistant Staphylococcus aureus* (MRSA).(31)

The findings suggest that gaming may be associated with a 9% higher score. This would be often be sufficient to avoid the severe consequences of an adverse inspection report, ranging from warnings to enforcement action of even restrictions on activity, with implications for the tenure of senior executives.

Our findings have obvious implications for policy, given the importance of hospital cleanliness in the fight against anti-microbial resistance (AMR). However, they also have implications for systems of regulation and inspection. One obvious question is whether inspections should be announced or unannounced. This has several implications. For example, our findings suggest that hospitals invested considerable resources on preparing for an inspection. On the other hand, it is arguable that they should be investing those resources at all times. A recent systematic review found only 3 studies.(32) The authors concluded that unannounced

inspections reduce the regulatory burden but there was no significant difference in what they found.

A further question is the extent to which a system based on inspections is the best way of ensuring quality. A history of regulation in the NHS in England described a series of shifts from trust-based professional regulation to detailed external inspection, followed by some rolling back.(33) Changes were often driven by events that questioned the system in place at the time, rather than evidence of clear superiority of an alternative.

While the characteristics of an ideal system are easy to specify, combining high standards with transparency, in practice, they seem more difficult to achieve. However, one lesson is clear. In any regulatory system it should be assumed that gaming will take place. The system should be designed in ways that minimise this.

References

1. Arrow KJ. Uncertainty and the welfare economics of medical care. The American economic review. 1963;53(5):941-73.

2. Williamson OE. Markets and hierarchies: some elementary considerations. American Economic Review. 1973;63(2):316-25.

3. Simborg DW. DRG creep: a new hospital-acquired disease. New England Journal of Medicine. 1981;304(26):1602-4.

4. House of Commons Committee of Public Accounts. Care Quality Commission. Twelfth report of session 2015-16. Dec 2015. www.publications.parliament.uk/pa/cm201516/cmselect/cmpubacc/501/501.pdf.2015.

5. Hussain I. School inspections: can we trust Ofsted reports? British Politics and Policy at LSE. 2012.

6. Figlio DN, Getzler LS. Accountability, ability and disability: Gaming the system? Advances in applied microeconomics. 2006;14:35-49.

7. Figlio DN. Testing, crime and punishment. Journal of Public Economics. 2006;90(4):837-51.

8. Jacob BA. Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. Journal of public Economics. 2005;89(5):761-96.

9. Figlio DN, Winicki J. Food for thought: the effects of school accountability plans on school nutrition. Journal of public Economics. 2005;89(2):381-94.

10. Bokhari FA, Schneider H. School accountability laws and the consumption of psychostimulants. Journal of Health Economics. 2011;30(2):355-72.

11. Anderson PM, Butcher KF, Schanzenbach DW. Adequate (or adipose?) yearly progress: Assessing the effect of" no child left behind" on children's obesity. National Bureau of Economic Research; 2011.

12. Washer P, Joffe H. The "hospital superbug": social representations of MRSA. Social science & medicine (1982). 2006;63(8):2141-52.

13. Daily Mail. One in 16 pick up a bug in FILTHY hospitals. http://wwwdailymailcouk/health/article-2606425/NICE-blames-staff-hygiene-dirty-equipment-thousands-deaths-One-16-pick-bug-FILTHY-hospitalshtml (accessed on 18th November 2016). 2014.

14. Daily Mail. Hospital bugs 'spread by use of wet wipes to clean wards. http://wwwdailymailcouk/news/article-3116162/Hospital-bugs-spread-use-wet-wipes-clean-wardshtml (accessed on 18th November, 2016). 2015.

15. Daily Mail. Hospital patient so shocked at dirty ward she climbed out of bed to clean it herself” 3rd july 2009. http://wwwdailymailcouk/news/article-1196997/Hospital-patient-shocked-dirty-ward-climbed-bed-clean-herselfhtml (accessed on 18th November, 2016). 2009.

16. Daily Mail. Vulnerable patients 'at high risk' of food poisoning. http://wwwdailymailcouk/wires/pa/article-3895752/Vulnerable-patients-high-risk-food-poisoninghtml (accessed on 18th November, 2016). 2016.

17. Daily Mail. Dirty hospital equipment causes 11 people to become infected with bacteria that could lead to PNEUMONIA. http://wwwdailymailcouk/health/article-3265031/Dirty-hospital-equipment-causes-11-people-infected-

bacteria-lead-PNEUMONIAhtml (accessed on 18th November 2016). 2015.

18. Department of Health. The NHS Plan: a plan for investment, a plan for reform. Department of Health, Publications and Policy guidelines (Access on line on 3rd March, 2016). 2000(http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/@ps/documents/digitalasset/dh_118522.pdf).

19. Mannion R, Braithwaite J. Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. Internal medicine journal. 2012;42(5):569-74.

20. Bevan G, Hood C. What's measured is what matters: targets and gaming in the English public health care system. Public administration. 2006;84(3):517-38.

21. Wismar M, McKee M, Ernst K, Srivastava D, Busse R. Health targets in Europe. Learning from experience European Union: World Health Organization. 2008.

22. Gravelle H, Sutton M, Ma A. Doctor Behaviour under a Pay for Performance Contract: Treating, Cheating and Case Finding?*. The Economic Journal. 2010;120(542):F129-F56.

23. Pickin M, Sampson F, Munro J, Nicholl J. General practitioners' reasons for removing patients from their lists:

postal survey in England and Wales. BMJ. 2001;322(7295):1158-9.

24. Care Quality Commission PIE. NHS Patient Survey Programme (Acute Trusts: Adult Inpatients Survey). In: National Health Service CQC, editor. Uk Data Service2010-2014.

25. Health & Social Care Information Centre. Patient Environment Assessment Team (PEAT). In: Health Do, editor. Health & Social Care Information Centre2010-2014.

26. Health & Social Information Centre. Patient-Led Assessments of the Care Environment (PLACE), England In: Health Do, editor.: Health & Social Information Centre; 2013-2014.

27. Health & Social Care Information Centre. Estates Return Information Collection. In: Health Do, editor. 2010-2014.

28. Toffolutti V, McKee M, Stuckler D. “Gaming the System”: Testing the Validity and Reliability of NHS Inspections of Hospital Cleanliness: Web Appendix 1. 2016.

29. Hahn J, Todd P, Van der Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica. 2001;69(1):201-9.

30. Card D, Krueger AB. Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. National Bureau of Economic Research; 1993.

31. Toffolutti V, Reeves A, McKee M, Stuckler D. Outsourcing cleaning services increases MRSA incidence: Evidence from 126 english acute trusts. Social science & medicine (1982). In Press.

32. Klerks MC, Ketelaars CA, Robben PB. Unannounced, compared with announced inspections: A systematic review and exploratory study in nursing homes. Health policy. 2013;111(3):311-9.

33. Bevan G. Changing paradigms of governance and regulation of quality of healthcare in England. Health, risk & society. 2008;10(1):85-101.

34. Mascle-Taylor HH, Taj; May, Ruth. A&E dalays: Why did patients wait longer last winter? . Monitor: Making the health sector work for patients (https://wwwgovuk/government/uploads/system/uploads/attachment_data/file/458764/AE_delay_main_report_finalpdf), accessed on 18/07/2016. 2015.

35. Johnson AP, Davies J, Guy R, Abernethy J, Sheridan E, Pearson A, et al. Mandatory surveillance of methicillin-resistant Staphylococcus aureus (MRSA) bacteraemia in England: the first 10 years. Journal of antimicrobial chemotherapy. 2012;67(4):802-9.

Figures and Tables

Exhibit 1: Descriptive Statistics, for 205 hospital sites between 2011 and 2014

Exhibit 2: Median perception of excellent cleanliness among 205 hospital sites between 2011 and 2014, by proximity to the month of assessment

Exhibit 3: Illustrative examples: Royal National Hospital for Rheumatic Diseases in 2013 and 2014

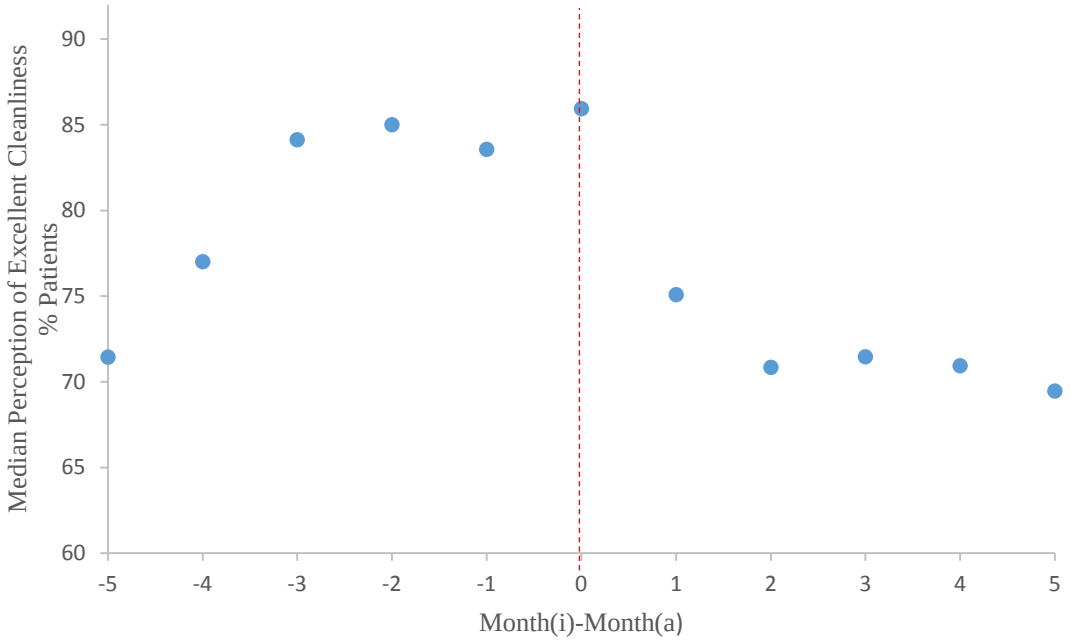
Exhibit 4: Change in the median percentage of patients' reporting excellent in audit months and in a short period previously

Exhibit 1: Descriptive Statistics, for 205 hospital sites between 2011 and 2014

Variable	Number of Hospital Months	Mean (S.D.)	Min.	Max.	Source
Median percentage of Patients reporting Excellent cleanliness	924	72.1% (11.4)	25	100	NHS Inpatients Survey
Number of Beds	913	637 (493)	5	2257	ERIC
Average length of stay in the Trust in days	924	6.07 (1.56)	2.40	14.2	NHS Inpatients Survey
Multiservice sites (dummy)	924	0.08 (0.27)	0	1	PEAT/PLACE
Specialist sites (dummy)	924	0.20 (0.40)	0	1	PEAT/PLACE
North of England (dummy)	924	0.44 (0.50)	0	1	ERIC
Central of England (dummy)	924	0.27 (0.44)	0	1	ERIC
London (dummy)	924	0.11 (0.32)	0	1	ERIC
South of England (dummy)	924	0.18 (0.39)	0	1	ERIC
Hospitals for each month of assessment	924	145 (49.9)	1	194	PEAT/PLACE
Number of patients with non-missing data on hospital cleanliness survey per month	924	205 (150)	1	552	PEAT/PLACE

Notes: Merged data at hospital level from Hospital data from Patient Environment Action Teams (PEAT) dataset (from 2011 till 2012), Patient-Led Assessments of the Care Environment (PLACE) (2013-2014), ERIC (Estates Return Information Collection) (2011-2014), NHS Inpatient Survey (2011-2014).

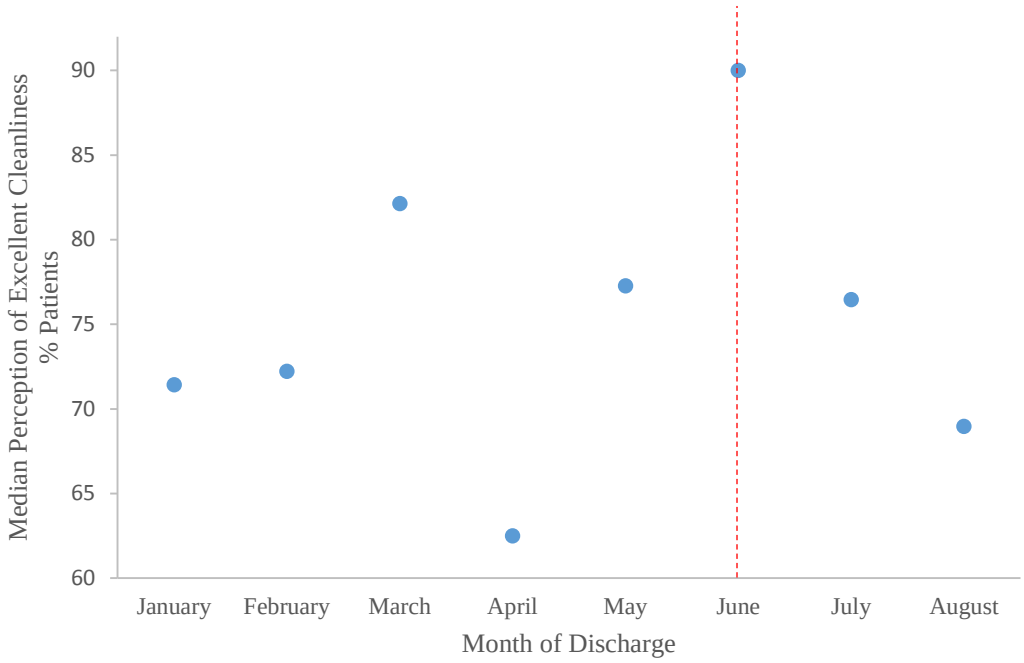
Exhibit 2: Median perception of excellent cleanliness among 205 hospital sites between 2011 and 2014, by proximity to the month of assessment



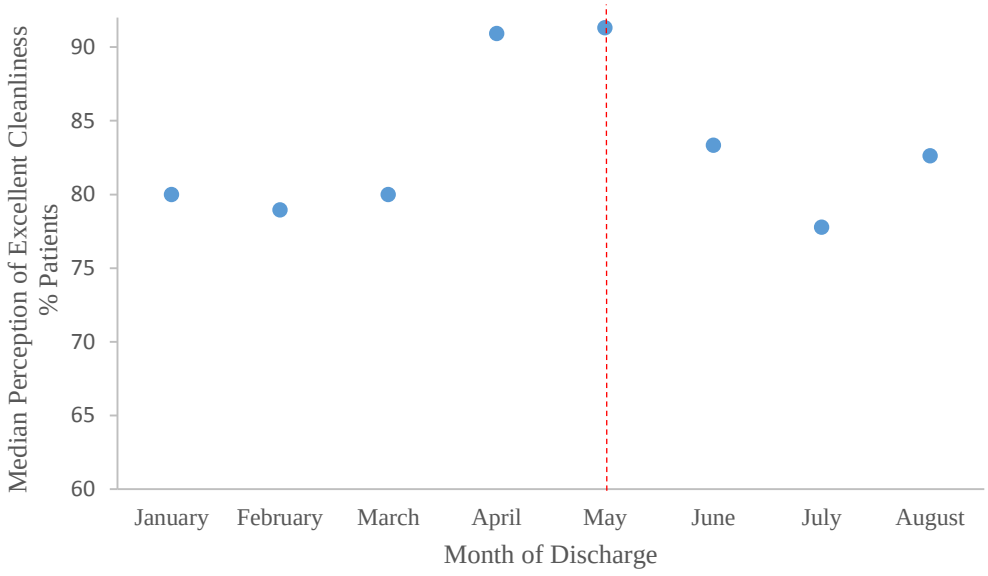
Notes: Merged data at hospital level from Hospital data from Patient Environment Action Teams (PEAT) dataset (from 2011 till 2012), Patient-Led Assessments of the Care Environment (PLACE) (2013-2014), ERIC (Estates Return Information Collection) (2011-2014), NHS Inpatient Survey (2011-2014). Dots represent the median percentage of patients reporting excellent cleanliness for the room or ward were they stayed in, red dash line represents the month of assessment. The number of hospitals used to compute the graph (at month level is): 1(5), 5(4), 4(3) , 5(2) 9(1), 18(0), 59(1), 131(2), 135 (3), 156 (4), 194(5). The average number of patients who responded to the questionnaire (at month level is): 14(5), 112(4), 29(3) , 30.6(2) 41.7(1), 113(0), 152(1), 189(2), 214 (3), 208(4), 231(5).

Exhibit 3: Cleanliness perception for Royal National Hospital for Rheumatic Diseases in 2013 and in 2014, by month of assessment. Dots represent the median cleanliness perception by month, the red-dashed line represents the month of assessment

a) Inspection in June 2013



b) Inspection in May 2014



Notes: Merged data for the Royal National Hospital Rheumatic Diseases Hospital form Patient-Led Assessments of the Care Environment (PLACE) (2013) and NHS Inpatient Survey (2013). A) The average number of patients who responded to the questionnaire (at month level is): 14(Jan), 18(Feb), 28(23), 16(Apr), 22 (34), 20 (Jun), 17 (35), 29 (Aug). B) The average number of patients who responded to the questionnaire (at month level is): 15(Jan), 19 (Feb), 10 (23), 11(Apr), 23 (34), 12 (Jun), 18 (35), 23 (Aug).

Exhibit 4: Change in the median percentage of patients' reporting excellent in audit months and in a short period previously

	Median patients' perception of excellent cleanliness Unadjusted Models	Median patients' perception of excellent cleanliness- Outsourcing hospitals vis-à-vis in House ones- Unadjusted	Median patients' perception of excellent cleanliness Fixed Effect Model	Median patients' perception of excellent cleanliness Fixed Effect Model
Inspection month	7.78%***	2.68%	2.54%**	2.86%**
	(2.55)	(3.14)	(1.27)	(1.42)
Time to inspection	-0.41	-0.46	0.89%***	1.04%***
	(0.59)	(0.54)	(0.29)	(0.36)
Time to inspection^2	-0.09	-0.08	-0.09%***	-0.02
	(0.07)	(0.07)	(0.04)	(0.08)
Time to inspection^3				-0.01%
				(0.1)
Outsourcing		-0.81		
		(1.91)		
Outsourcing * Inspection Month		11.0%***		
		(2.75)		
Size	No	No	Yes	Yes
Complexity	No	No	NA	NA

Non-linear Time-trend (including quarter dummies)	Yes	Yes	Yes	Yes
R^2	0.05	0.06	0.04	0.04
Number of hospital-months	907	867	907	907
Number of hospital	205	200	205	205

Notes: Source: Data from Hospital data from Patient Environment Action Teams (PEAT) dataset (from 2011 till 2012), Patient-Led Assessments of the Care Environment (PLACE) (2013-2014), ERIC (Estates Return Information Collection) (2011-2014), NHS Inpatient Survey (2011-2014), NHS Staff Survey (2011-2014), and Public Health for England (2011-2014). SE clustered at hospital site level.
The dependent variable represents the median patients' perception of excellent cleanliness for the room or ward where they stayed. The dependent variable has been aligned, through regression, on a quadratic function of the distance to the month of assessment, the assessment dummy, the number of beds (size), whether the site is a specialist or a multiservice or another site-type (complexity), year and quarter dummies (non-linear time trend).
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$