Management Practices in Hospitals

Nicholas Bloom

Stanford University, NBER and Centre for Economic Performance

Carol Propper

Imperial College and CMPO

Stephan Seiler

London School of Economics, Centre for Economic Performance

John Van Reenen

London School of Economics, Centre for Economic Performance, NBER and CEPR

This draft: December 29th 2008

Preliminary

Abstract

We develop a new methodology for measuring management practices in hospitals, and use this in 182 interviews of physicians and managers in public and private hospitals (covering 61% of English acute trusts). We find our management measure is strongly correlated with hospital performance, both clinical outcomes like survival rates from heart attacks, and general operational and financial outcomes. Management in publicly owned hospitals (the National Health Service) compares poorly with management in manufacturing. These public hospitals also appear to have significantly worse management practices than private hospitals. Among publicly owned hospitals management scores are relatively higher for Foundation Trusts (hospitals with greater autonomy from the government), for larger hospitals and where managers have more clinical expertise. We also find some evidence that competition is associated with better hospital performance.

JEL classification: J45, F12, I18, J31

Keywords: management, hospitals, competition, productivity

Acknowledgements: We would like to thank Pedro Castro, John Dowdy, Stephen Dorgan, Ben Richardson for discussions. We thank the managers and physicians who took part in the survey. Particular thanks are due to our team of interviewers. Financial support is from the ESRC through the Centre for Economic Performance and the EDS Innovation Centre.

Corresponding author: j.vanreenen@lse.ac.uk; Centre for Economic Performance, LSE, Houghton Street, London, WC1E 2AE, UK.

All over the world, healthcare costs are rising as a proportion of national income. In the UK, for example, healthcare rose from 7.1% of GDP in 2001 to 9.4% in 2006, while in the US this has risen from 14.5% to 16% over the same period, with both projected to rise further (Hall and Jones, 2007). Escalating costs has led to a much greater emphasis on improving productivity in healthcare, especially since a large share of these costs are subsidised by the government.

We know that there are large differences in hospital performance across a wide range of indicators even after extensive controls have been made for differential case mix and hospital inputs (Kessler and McLennan, 2000; Hall et al, 2008). This is not so surprising – there is a huge variability in productivity in many other areas of the private and public sector (e.g. Foster, Haltiwanger and Syverson, 2008). Commentators have long believed that these performance differences were at least in part linked to management practices, but the main evidence for this belief resides in anecdote and from case studies rather than systematic quantitative evidence.

In recent work we have pioneered a methodology for quantifying management practices and implemented this survey tool on thousands of manufacturing firms in Europe, Asia and the US (Bloom and Van Reenen, 2007; Bloom et al, 2007). The measures proved very robust to measurement error and our management scores were strongly correlated with firm performance. The manufacturing sector is a declining share of employment and GDP for developed nations, however, so a legitimate question is whether the survey tool can also be used in other sectors. In this paper we apply the same basic methodology to measuring management in the healthcare sector. We implement our methods in 161 interviews across 100 English acute hospital trusts interviewing a mixture of clinicians and managers in two specialities: cardiology and orthopaedics. On top of that, 21 private sector hospitals were also interviewed using the same methodology. We cover 61% of all providers of acute care in the UK.

Our results are both methodological and substantive. On the methodological front, we show that our management practice scores deliver useful information and are correlated with measures of hospital performance such as lower mortality rates from AMI[1] and general surgery, waiting lists, staff turnover and composite measures of performance. On the substantive front we uncover several interesting findings:

[1] Acute myocardial infarction, commonly known as a “heart attack”.

First, the average scores of management are lower in hospitals than for manufacturing. This is primarily due to much different "people" management which includes hiring, firing, promotions, rewards and recruitment. Targets are also a problem in the NHS with many being arbitrarily imposed from central government. Second, the average scores of management are lower in the public than in private hospitals, with again this gap primarily due to people management. These differences between government and non-government hospitals are consistent with Duggan (2000) who finds large differences in behaviour of these hospital types in US data[2].

Third, we find that when managers have clinical qualifications, average management scores are significantly higher. This suggests that the asymmetry of information between managers and the powerful interests of senior doctors is a key factor that leads to lower performance.

Finally, we find some evidence that competition is associated with better hospital performance. This effect is smaller than the comparable results for private sector manufacturing, suggesting competitive forces are more constrained in healthcare. This inhibits the exit or takeover of poorly performing hospitals.

The structure of the paper is as follows. The next section discusses the data, Section II describes the relationship between performance and management and Section III contrasts public healthcare with private healthcare and other sectors in the UK and internationally. Section IV describes the factors that are strongly associated with management in the public health sector. Section V offers some concluding comments.

I. DATA

The data used for the analysis is drawn from three different sources: the management survey conducted by the Centre for Economic Performance at the London School of Economics, which includes 18 questions from which the overall management source is computed plus additional information about the process of the interview and features of the hospitals. This is complemented

[2] Duggan (2000) shows that for-profit and not for profit hospitals behaved in a similar way when faced with a large change in financial incentives to treat low income patients (i.e. they were much more responsive than government hospitals and tended to cream skim the easier to treat, but poorer, patients). This is consistent with the survey in Sloan (2000).

by external data from the UK Department of Health, which provides information on many hospital characteristics such as clinical outcomes, patient case mix, size and measures relating to the quality and efficiency of treatment.

I.A. Management Survey Data

The core of this dataset is made up of 18 questions which can be grouped in the following three subcategories: operations (3 questions), monitoring (3 questions), targets (5 questions) and people management (7 questions). For each one of the questions the interviewer reports a score between 1 and 5, a higher score indicating a better performance in the particular category. Table B2 shows descriptive statistics for all individual questions and averages for the subcategories. The last two columns report the equivalent score from the manufacturing sample and the difference between the average scores for manufacturing firms and hospitals.[3] A detailed description of the individual questions and the scoring method is provided in Appendix A.[4]

A key challenge in evaluating these management questions is to obtain unbiased responses. To try to do this we used a *double-blind* survey methodology. The first part of this was that the interview was conducted by telephone without telling the respondents that they were being scored. This enabled scoring to be based on the interviewer's evaluation of the hospital's actual practices, rather than their aspirations, the respondent's perceptions or the interviewer's impressions. To run this "blind" scoring we used open questions (i.e. "can you tell me how you promote your employees"), rather than closed questions (i.e. "do you promote your employees on tenure [yes/no]?"). Furthermore, these questions target actual practices and examples, with the discussion continuing until the interviewer can make an accurate assessment of the hospital's typical practices based on these examples. For each practice, the first question is broad with detailed follow-up questions to fine-tune the scoring. For example, in dimension (1) *Layout of patient flow* the initial question is "Can you briefly describe the patient journey or flow for a typical episode?" is followed up by questions like "How closely located are wards, theatres, diagnostics centres and consumables?"

[3] There are 16 questions in the manufacturing survey, which overlap with the hospital survey. Therefore the comparison is only possible for these 16 questions. The manufacturing sample includes all firms based in the UK, including multinationals.

[4] The questions in appendix A correspond in the following way to these categories. Operations: question 1-3, Monitoring: question 4-6, Targets: question 8-12, People management: question 7 and 13-18.

The second part of the *double-blind* scoring methodology was that the interviewers did not know anything about the hospital's performance in advance of the interview. The interviewers were specially trained graduate students from top European and U.S. business schools. Since each interviewer also ran 46 interviews on average we can also remove interviewer fixed effects in the regression analysis.

The survey also includes questions on other features of the hospital such as the number of sites and the number of managers with a clinical or managerial degree. Whenever these variables can more reliably be obtained from the external dataset (see below) we cross check results against this source as well.

Finally, we also collected a set of variables that describe the process of the interview, which can be used as "noise controls" in the econometric analysis. The variables collected included: an interviewer fixed effect, the time of the day and date of the interview, the duration of the interview, the position of the interviewee (clinician or manager), the speciality in which he is located (cardiology or orthopaedics) and a reliability index coded by the interviewer. The interviewee's tenure in the post and in the trust is also reported. Including these "noise controls" helps reduce residual variation.

Obtaining interviews with managers was facilitated by the endorsement of the Department of Health, and the name of the London School of Economics, which is well known in the UK as an independent research university. This strong government and academic endorsement enabled us to interview respondents for an average of just under an hour.

I.B. External Data

In the manufacturing sector economists generally use labour or total factor productivity as a measure of organizational performance. In the case of hospitals it is more difficult to measure output, particularly where patients do not pay directly for their care and standard productivity measures are therefore not available. It is not straightforward to develop a single summary measure of hospital performance and data restrictions limit the indicators that are available on a consistent cross-hospital basis. As the main goal of hospitals is to improve its patients' health, variables capturing the success of treatment such as mortality rates are a natural candidate. Another possibility is to use a broader measure that also takes financial efficiency, resource use and other factors into account. Hospital regulators in the USA and the UK use a wide range of measures in their attempts to assess hospital performance[5]. The sources of these are detailed in Appendix Table B1.

We therefore examine the correlation of each of a number of clinical and non-clinical performance measures with the management score. The key clinical outcomes we use are the 28 day mortality rate for non-elective (i.e. emergency) admissions for (i) AMI (acute myocardial infarction)[6] and (ii) non-elective surgery[7]. We choose these for three reasons. First, regulators in both the USA and the UK use selected death rates as part of a broader set of measures of hospital quality. Second, using emergency admissions helps to reduce selection bias because elective cases may be non-randomly sorted towards hospitals. Third, death rates are well recorded and cannot be "gamed" by administrators trying to hit targets. Fourth, heart attacks and overall emergency surgery are the two most common reasons for admissions that lead to deaths.

As another performance indicator we use the size of the waiting list for all operations. Long waits have been an endemic problem of the UK NHS; although these have fallen dramatically over the last 8 years (see Propper et al, 2008). We also use MRSA infection rates ("superbugs") as a further quality measure for the hospital. Again, both of these measures have been used by the UK government to rate NHS hospitals.

These indicators have the disadvantage that each individual measure is rather noisy so aggregating into a summary hospital performance score is desirable. There is an element of subjectivity in deciding what set of performance metrics to use and what weight to put on each individual metric. To avoid any concern that we are choosing these arbitrarily, we use the Department of Health's own Health Care Commission ratings which represent such a composite performance measure. The Health Care Commission's rates hospitals along two dimensions of "resource use" and "quality of service" (measured on a scale from 1 to 4). The efficiency of resource use is measured by the number of spells per medical employee, bed occupancy rate and the average length of stay. Service

[5] See for example http://2008ratings.healthcarecommission.org.uk/informationabouthealthcareservices.cfm
[6] Examples of the use of AMI death rates to proxy hospital quality include Kessler and McClellan (2000), Gaynor (2004) and, for the UK, Propper et al (2004).
[7] Death rates following emergency admission were used by the UK regulator responsible for health quality in 2001/2 .2001/2 CHI indicators for 2001/2 http://www.performance.doh.gov.uk/performanceratings/2002/tech_index_trusts.html

quality is measured by clinical outcomes (readmission risk and infection rates), waiting times and a measure of patient satisfaction as well as job satisfaction of the staff. We use the 2006 values as these are coincident with the timing of the survey and average across the two measures (which are on a scale of 1 to 4). These ratings replaced the HCC's single "star rating" (on a scale of 1 to 3). The HCC does not reveal the exact formula it uses to aggregate over the components of the index, but some averaging is valuable due to the noisiness of the underlying performance measures. We also report experiments where we disaggregate the index and construct our own (re-aggregated) index.

We also collected data on total employment, the number of doctors, beds, speciality, location etc. as additional control variables. The descriptive statistics for some of the most important variables, which will be used later on, can be found in Table 1. The mortality rate from AMI is 17%, although there is considerable variation (e.g. Hall et al, 2008) whereas it is lower for surgery. A typical hospital trust has 3,651 staff, 387 medical full-time equivalents (physicians) and 15,513 patient-cases per quarter. These may seem large because a typical trust is multi-site (2.6 on average).

I.C Descriptive Statistics

We approached up to four individuals in every hospital – a manager and physician in the cardiology service and a manager and clinician in the orthopaedic service. There were 164 acute hospital trusts with orthopaedics or cardiology departments in England and 61% of hospitals (100) responded which is a very high hit rate for a voluntary survey. We obtained 161 interviews, 79% of which were with managers (it was harder to obtain interviews with physicians). The responses between the two service lines were evenly split. Furthermore, we show that response probability was uncorrelated with observables such as performance outcomes, size and composition (Appendix B). We also ran a smaller scale survey asking identical questions private hospitals and collected information on 21 of these. Again, we could find no systematic response bias, although the number of observables for private hospitals is much smaller.

I.D. Preliminary Data Analysis

Before any econometric estimation we first present some simple descriptive statistics. In Figure 1 we present the non-parametric plot of the relationship between the HCC average rating and the management practice score. There appears to be a positive correlation between the two variables, suggesting that the management responses are not simply "cheap talk". Figure 2 presents a similar graph, cut slightly differently. We divide the HCC score into quintiles and show the average management score in each bin. There is a clear upward sloping relationship: management scores in the lowest quintiles are 2.3 and 2.4., in the next two quintiles they are between 2.5 and 2.6 and in the highest quintile they are 2.8.

Figure 3 plots the entire distribution of management scores for our respondents (in the upper Panel A). There is a large variance with some well managed firms, and other very poorly managed. It is striking that there are few hospitals which scores above a 4. In Panel B we present a comparison between hospitals and UK manufacturing firms. To make the samples somewhat comparable we keep only establishments who have between 50 and 5000 employees and who are domestically owned (i.e. we drop multinationals from the manufacturing sample). Furthermore, in both panels we are using the average management score from only the comparable 16 questions, because two questions on lean management are difficult to compare across sectors.[8] Hospitals clearly have lower management scores than manufacturing firms. Table B2 shows that this is particularly true of people management and targets. We will investigate this in more detail in Section III below.

II HOSPITAL PERFORMANCE AND MANAGEMENT PRACTICES

Before examining the factors "driving" management practices we will first check that the management score is robustly correlated with external performance measures. This is not supposed to imply any kind of causality. Instead, it merely serves as an "external validity" check to see whether a higher management score is correlated with a better performance.

We estimate regressions of the form:

$$y_i^k = \alpha M_{ij} + \beta' x_{ij} + u_{ij}$$

Where y_i^k is performance outcome k (e.g. AMI mortality) in hospital i. M_{ij} is the average management score of respondent j in hospital i, x_{ij} is a vector of controls and u_{ij} the error term. Since errors are correlated across respondents within hospitals we cluster our standard errors at the hospital level (they are also robust to heteroscedacity)[9]. We present the performance and management measures in z-scores so the tables can be read as the association of a one standard

[8] The questions we dropped are 1 and 2 in Appendix A.
[9] Furthermore we weight the observations with the inverse of the number of interviews conducted at each hospital. This gives equal weight to each hospital in the regressions.

deviation of management on the outcome (all results are robust to this normalization). We consider disaggregating the 18 questions below, but our standard results simply z-score each individual question, average these into a composite and then z-score this average. In terms of timing, we use the 2005/6 average outcomes in the year to be consistent with the management survey[10].

An important control for the outcomes is the casemix of the patients. We use casemix adjustments standard for the clinical condition we examine. We have the age/gender profile[11] of all admissions for each type of condition (e.g. the demographic profile of patients admitted with AMI in hospital *i* in a given year). In all regressions we also control for the mortality rates in the hospital's "catchment area" to reflect the fact that worse outcomes are likely if the hospital is located in a community with a high rate of ill health (e.g. many old people or high poverty rates).

The other control variables can be grouped into "general controls" and "noise controls". The general controls contain regional dummies (10), a dummy for whether the speciality is in cardiology or orthopaedics, a size proxy (the total number of patient cases at the hospital level)[12]. Noise controls comprise interviewer dummies (4), interview characteristics (duration of the interview and the number of management questions not answered) and interviewee characteristics (tenure, whether the respondent was a clinician or manager).[13]

Table 2 shows results for regressions of each of the performance measured on the standardized management score. The management score is the top panel calculated as the average of 16 out of the 18 questions in the survey excluding the operations questions.[14] The bottom panel shows results based on all 18 questions. The first thing to note, looking at the first row of the table is that higher management scores are associated with better hospital outcomes across all the measures and this

---

[10] We also used longer time averages going back to 2001 in an effort to assess the importance of transitory measurement error. The qualitative results were similar, but actually tended to weaken as we used years further away from the date of the management survey.
[11] Specifically we have 11 age categories for each gender (0-15, 16-45, 46-50, 51-55, 56-60, 61-65, 66-70, 71-75, 76-80, 81-85, >85), so up to 22 controls. These are specific to the conditions (AMI, surgery, etc.) considered. For the general performance indicators (like HCC rating) we use all patients admitted.
[12] We also experimented with a number of other size controls such as total employment, the number of sites in the trust, the number of acute beds and the number of medical FTEs. These gave similar results to using patient-cases.
[13] In order to avoid losing many observations whenever a control variable was missing, we replace the missing value with the mean value of the variable and generate a dummy variable equal to unity for the missing observation. This is included in the regression together with the modified original variable. The results are robust to dropping the missing values.
[14] In Figure 3 and Table 3 we use a pooled sample of the hospital and a manufacturing sector survey. The two surveys are comparable for all but the two operations questions, so these have to be excluded.

relationship is significant in every case except for MRSA infection rates. This immediately suggests our measure of management is not simply cheap talk, but has informational content.

In the first column of Table 2 the AMI mortality rate is regressed on the management score controlling for a wide number of confounding influences[15]. As is standard we drop observations where the number of cases admitted for AMI is low because this leads to large swings in observed mortality rates[16]. High management scores are associated with significantly lower mortality rates from AMI: a one standard deviation increase in the management score is associated with just under 0.1 of a standard deviation fall in AMI mortality. Columns (2) and (3) examine death rates from different types of surgery (the second column is all emergency surgery and the third column is a sub-set of more highly risky). In both cases there is a significant correlation, although the point estimate is larger in column (3). Columns (4) and (5) use waiting list indicators as measures of poor hospital performance. These are not directly health outcomes, but they are closely related, as they measure how long it takes to receive a potentially health improving treatment. Better managed hospitals tend to have significantly lower waiting lists. Column (6) uses MRSA infection rates are used as an indicator of health outcomes (something that has been a government priority in recent years). The coefficient is correctly signed but insignificant.

A concern with the management measures is that they might be associated with higher efficiency at the expense of worse work quality. We use data from the NHS Staff Survey which asks all employees whether they intended to leave the hospital in the next year. We use the average of this measure across all workers in the hospital in column (7) as another performance outcome. Higher management scores are associated with a lower probability of wanting to exit the hospital.

The final columns use a rating by the Health Care Commission (HCC) of UK hospitals. We average the HCC's rating on "resource use" and "quality of service" in column (8). We also compute a "pseudo HCC rating" by attempting to reverse engineer the process by which the original rating was calculated (see Data Appendix B) in column (9). The management practice score is significantly and positively correlated with both of these measures. When using the individual components of the "pseudo-rating" as dependent variables in the regression although the coefficient on management is

[15] Controlling for case mix is particularly important. Without controls for casemix the coefficient is positive and insignificant. This suggests the better managed hospitals are actually taking on more of the complex high risk cases.
[16] Following Hall, Propper and Van Reenen (2008) we drop hospitals with under 150 cases of AMI. The results are not sensitive to the exact threshold.

always of the “correct” sign, only two components are significant at the 5% level: waiting times and staff job satisfaction. Averaging over different outcome variables increases the significance of the right hand side variables which suggests that averaging helps mitigate measurement error[17].

The lower row of Table 2 repeats the exercise over all 18 questions with very similar results. Overall, the Table 2 is reassuring in that our measure of management practices is associated with superior hospital outcomes across a wide range of performance indicators[18].

III COMPARING MANAGEMENT PRACTICES ACROSS SECTORS

As a next step we compare management practices in the healthcare sector with management in manufacturing firms. We use data from the equivalent survey of management practices in the manufacturing sector (see Bloom and Van Reenen, 2007; Bloom et al, 2007). In order to make the two datasets comparable we only use 16 out of 18 questions. Thus, we have a large manufacturing sample of around 651 firms and 182 hospital interviews (including 21 private sector hospitals).

In column (1) of Table 3 we simply regress the management practices score on a dummy for being a hospital, with the manufacturing as the baseline. As suggested in Figure 3 hospitals appear to score significantly worse on management than manufacturing (about half of a standard deviation). In column (2) we add a dummy for privately owned hospitals, which is positive and highly significant. The coefficient on the hospitals dummy becomes more negative in this specification, as the higher management practice score of private hospitals is now separated from the public ones. This indicates our sample of private hospitals scores more highly than manufacturing firms which are also all privately owned in the UK. In column (3) we replicate column (2) and include a control variable for the size of the hospital or firm (total employment in the hospital or in the firm). The number of observations is reduced as we do not have information on employment for the whole sample (e.g.

---

[17] We also examined decomposing the management score. When regressing them individually on the HCC rating 11 questions out of 18 questions are significant at the 10% level (11 of them are significant at the 5% level and 4 are significant at the 1% level). When regressing the averages of the four subcategories operations, monitoring, targets and people management individually on the HCC rating we obtain significant coefficients at the 1% level in all cases but the operations category. If all four variables are regressed on the HCC rating only the incentive questions are significant (at the 10% level).

[18] We also looked at the effect of the different subcategories of the management score (operations, monitoring, targets and people). The management score based only on the subset questions belonging to a particular category was regressed on different health outcomes using the same regressions as above. Overall “target” and “people” questions have the most explanatory power for the different health outcomes followed closely by the “monitoring” category of questions.

privately owned hospitals). Larger organizations tend to have higher managerial quality (see Lucas, 1978), but the magnitude and significant of the other coefficients is little altered.

The differences between the NHS and private hospitals could arise from many factors. One possibility is that the mix of treatments is very different as UK private hospitals specialize in elective treatments for which there are long waiting lists in the NHS – they do not have to maintain emergency rooms that must by law accept all patients irrespective of their ability to pay. This may make them intrinsically easier to manage. An alternative explanation is that government control may place many constraints over the ability of hospitals to be effectively managed. We try to shed some light in this in two ways. First, we disaggregate the management questions by sub-groups of “types” and second we look at government controlled firms in the manufacturing sector in other countries.

In columns (4) to (6) we look at the management scores for subcategories of the 16 questions. In column (4) we start by looking at *monitoring* management, which covers questions 4 – 6 in Appendix A, focusing on the collection and use of information. We see that NHS hospitals score significantly lower than manufacturers at monitoring management practices and private hospitals perform significantly better than public ones. In column (5) we find very similar results for the *targets* category (questions 8 - 12). The difference both between hospitals and manufacturing and between public and private hospitals is more pronounced for this category of questions. Finally, when looking at *people management* in column (6), which cover questions 7 and 13 - 18, focusing on hiring, firing, pay and promotions management, we also obtain a negative and significant coefficient for the hospital dummy term. Also, private hospitals again score more highly than public ones. The coefficient on the private hospital dummy is positive and significant and larger than for the other two categories. The low score for NHS hospitals on targets may reflect the fact that there are a huge number of detailed and often mutually inconsistent targets that are handed down to NHS hospitals from the Centre (“Command and Control”). The low scores on people management may reflect the high degree of central regulation and union power over hiring, pay and promotions.

In columns (7) - (9) we widen the sample still further using data on manufacturing firms from other European countries[19]. We do this in order to show a contrast between government and non-government owned (“private”) firms in the manufacturing sector as a whole (this cannot be done just

[19] See Bloom, Sadun and Van Reenen (2008) for a discussion of this larger survey.

for the UK as there are no government owned manufacturing firms in our sample). Column (7) simply includes a dummy for hospitals as in column (1) and shows a large negative coefficient as before. Column (8) also includes a private sector dummy and illustrates that privately owned firms score more highly on the management score than state owned firms (see also Bloom et al, 2007). The final column repeats our earlier specification on the UK which includes a dummy for private hospitals but also includes the private sector dummy from the previous column. The public-private difference in healthcare partly reflects a general public-private difference in management scores elsewhere in the economy. But the difference in healthcare is even stronger than that elsewhere (as a test of the difference between the management score of a private hospital and a private manufacturing firm has a p-value of 0.06).

In summary, publicly owned hospitals have a lower management score both compared to the manufacturing sector and with private sector hospitals. These are purely descriptive results and should not be read to say that the low scores of NHS hospitals are necessarily because they are publicly owned. Nevertheless, the pattern of results does suggest that the lack of autonomy of local managers in the centralized healthcare system of the UK may be behind the low management scores. We now turn to a deeper investigation of this.

IV EXPLAINING HEALTHCARE MANAGEMENT SCORES IN THE PUBLICLY OWNED HOSPITAL SECTOR

IV.A Autonomy, skills and size

To investigate the factors that influence hospital management score we regress the management score (for the 16 question used previously) several potentially relevant factors. These include a dummy variable for whether the hospital is a Foundation trust (a public sector hospital with greater autonomy from the Government), the number of medical employees, the proportion of doctors, the proportion of managers with a clinical degree and regional dummies. We also include the general controls and noise controls as in previous tables.

The results are presented in Table 4 with column (5) being our preferred specification which includes all covariates. In column (1) we see that Foundation trusts score more highly on management. This is an interesting result and accords with intuition that greater freedom from the

Government is associated with improved management practices[20]. Column (2) shows that the proportion of managers with a clinical degree is positive and significant in almost all specifications. This indicates that a separation of clinical and managerial knowledge inside the hospital is associated with worse management which may indicate that managers need to have some clinical knowledge in order to effectively challenge senior doctors. Interestingly, it is much rarer in the UK than the US for senior physicians to going into a senior managerial position such as Chief Executive of a large general hospital (the salaries and status of these positions is relatively less attractive in the UK).

Thirdly, there is some evidence that size, measured as the total number of patient cases at the hospital level is positively correlated with management scores. Although this is insignificant in column (3), it is significant when all the additional covariates are included in the final column[21]. The positive correlation of size and management was also revealed in the manufacturing sector. Since hospital size is not really influenced by performance, as there is little patient choice in the NHS, it is likely that larger hospitals are able to attract better quality managers.

IV.B Competition

Given the extensive discussion surrounding competition and performance in healthcare and other sectors (e.g. Kessler and McClellan, 2000; Nickell, 1996), we analyzed different measures for the intensity of competition. We begin by use the HCC rating of Table 2 column (8) as the dependent variable which averages across a number of desirable hospital performance indicators. Column (1) shows that the average hospital performance was positively and significantly correlated with the number of competing hospitals in the local area (defined as a 30km radius around the hospital)[22]. Column (2) includes the control for management practice scores which as expected enters with a positive and significant sign. The coefficient on competition falls because competition and management practices are positively correlated. This is shown by column (3) where we report

[20] We should note, however, that this result would also arise if Foundation trust status was only possible for hospitals that had better management. Although our scores were not used for this purpose we know that they are correlated with HCC rating, which was a factor.

[21] The smaller hospitals tend to have more managers with clinical training which is why omitting this variable causes an under-estimation of the size effect.

[22] We use all hospitals, but obtain similar results if define rivals as only public hospitals as their location is given by long-standing historical factors with very little exit or entry. We also examined other competition measures such as wider “markets” than 30km and the manager’s perceived level of competition. These were highly correlated with this measure and led to similar findings so we do not report the results.

regressions of the management practice score on the number of competing local hospitals. We find a positive coefficient on management which is significant at the 10% level.

Following recent US work[23] we also examined whether there are different responses by ownership type. We found that correlation of performance and competition was significantly stronger for private hospitals than public hospitals (p-value = 0.079). The differences in behavioural response between government and private hospitals reported here are similar in flavor to the findings in Duggan (2000).

An interpretation of Table 5 is that competition plays a role in improving hospital performance and this is partly through improving management practices (the coefficient on competition falls from 0.043 to 0.035 in column (2)). Competition does not seem as strong an influence on management in healthcare as manufacturing where similar regressions to column (3) yield much stronger relationships (see Bloom and Van Reenen, 2007). One reason why higher competitive pressure may not have so a strong impact on management practices is that autonomy is lower and the exit of underperforming hospitals is rare (due to political pressure). For private hospitals, competition (from both public hospitals and other private hospitals) is much more salient as exit is credible and they have greater autonomy to respond. Of course, things may be also be changing as reforms introduced in recent years have introduced more quasi-market elements to the British healthcare sector[24].

V CONCLUSIONS

In this paper we have described a new methodology for quantifying the quality of management practices in the healthcare sector. We have implemented this survey tool on almost two thirds of acute hospitals in England. We found that our measure of management quality was robustly associated with better hospital outcomes across mortality rates and other indicators of hospital performance. This is consistent with Bloom and Van Reenen's (2007) work in the manufacturing sector.

[23] See Cutler and Horwit (1999), Silverman and Skinner (2001) and Duggan (2002).
[24] We could not find evidence that Foundation trusts responded differentially, however. This may be because the reforms were still in their early stages in 2006.

Management in public hospitals scores significantly worse than firms in the manufacturing sector. Public hospitals also do worse managed than private hospitals, although the latter deal with a much smaller fraction of (wealthier) patients with less acute treatments. Among public sector hospitals management scores are significantly higher for Foundation trusts (hospitals with greater operating autonomy), for larger hospitals and where managers have more clinical expertise. We also find some evidence that product market competition is associated with better hospital performance and management.

In terms of future work, it would be extremely interesting to expand our sample to look at healthcare management in other countries. We have piloted some work along these lines and plan to implement this in the US and other nations. We also intend to look more closely at the role of competition exploiting changes in UK policy over recent years. Finally, examining how hospitals of different management quality and ownership respond differentially to shocks could be very revealing (Duggan, 2000).

REFERENCES

Bloom, Nicholas and Van Reenen, John (2007) "Measuring and Explaining Management practices across firms and nations", *Quarterly Journal of Economics*, Vol. 122, No. 4: 1351–1408.

Bloom, Nicholas, Dorgan, Stephen, Dowdy, John and Van Reenen, John (2007) "Management Practices Across Firms and Nations", LSE/McKinsey http://cep.lse.ac.uk/management/default.asp

Bloom, Nicholas, Sadun, Raffaella and Van Reenen. John (2008) "Measuring and explaining organizational practices across firms and nations" LSE/Stanford mimeo

Cutler, David M. and Jill R. Horwitz, "Converting Hospitals from Not-For-Profit to For-Profit Status: Why and What Effects?" in D. Cutler (ed.) *The Changing Hospital Industry: Comparing Not-For-Profit and For-Profit Institutions*, Chicago: University of Chicago Press and NBER (1999

Duggan, Mark (2000) "Hospital Ownership and Public Medical Spending" *Quarterly Journal of Economics*, November 1243-1374

Duggan, Mark (2002) Hospital Market Structure and the Behavior of Not-For-Profit Hospitals *The RAND Journal of Economics*, Vol. 33, No. 3 (Autumn, 2002), pp. 433-446

Foster, Lucia, John Haltiwanger, and Chad Syverson (2008) "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98(1), 394-425

Gaynor, Martin (2004) "Competition and Quality in Health Care Markets. What Do We Know? What Don't We Know?" *Economie Publique* 15: 3-40.

Hall, Emma, Propper, Carol and Van Reenen, John (2008) "Can Pay Regulation Kill? Panel Data evidence on the effect of labor markets on hospital performance", NBER Working Paper No. 13776

Hall, Robert and Jones, Chad (2007) "The Value of Life and the Rise in Health Spending" *Quarterly Journal of Economics*, 122(1), 39-72

Healthcare Commission (2006) "The annual Health Check in 2006/2007". http://www.healthcarecommission.org.uk/_db/_documents/The_annual_health_check_in_2006_2007_assessing_and_rating_the_NHS_200609225143.pdf

Kessler, Daniel P. and Mark B. McClellan (2000) "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics* 115 (May): 577-615.

Lucas, Robert (1978), "On the Size distribution of business firms", *Bell Journal of Economics*, IX (2), 508-523.

Nickell, Steve (1996), "Competition and Corporate Performance", *Journal of Political Economy*, CIV (4), 724-746.

Propper, Carol, Simon Burgess, and Katharine Green (2004) “Does Competition between hospitals improve the quality of care? Hospital death rates and the NHS internal market.” *Journal of Public Economics* 88 (July): 1247-1272

Propper, Carol, Matt Sutton, Carolyn Whitnall, and Frank Windmeijer (2008) “Did targets and terror reduce waiting times for hospital care in England” *The BE Journal of Economic Analysis and Policy, Issue 8, 2, Paper 5* (2008).
Silverman, E. and Skinner, J., (2001) “Are For-Profit Hospitals Really Different? Medicare ‘Upcoding’ and Market Structure,” *NBER Working paper No. W8133*

Sloan, F (2000) “Not-for-profit ownership and hospital behaviour” in Culyer, A J and Newhouse, J (eds) *Handbook of Health Economics*, pp 1141-1174. Elsevier: Amersterdam.

Table 1: Means and Standard Deviations of Variables

Variable	Mean	Standard Deviation
Mortality from emergency AMI after 28 days (quarterly average)	16.9	11.0
Mortality from all emergency surgery after 30 days (quarterly average)	2.56	1.11
Mortality from selected emergency general surgery after 30 days (admissions into General surgery Unit only, quarterly average)	5.20	1.70
Infection rate of MRSA per 10,000 bed days (half yearly)	1.57	0.89
Numbers on waiting list	5,764	3,226
Percentage on waiting list at risk of breaching national target	0.923	1.025
Likelihood of leaving in next 12 months (1=very unlikely, 5=very likely)	2.667	0.128
Average Health Care Commission rating (1-4 scale)	2.25	0.68
Pseudo HCC rating (standardized)	0.00	0.45
Proportion of physicians in total hospital employment	11	2
Managers with a clinical degree	50.3	31.7
Crude Mortality Rate in hospital's area (per 100,000 population)	932	138
Foundation Trust (hospitals with greater autonomy)	34.2	47.6
Number of competing hospitals in 30km radius (total)	20	27
Number of competing hospitals in 30km radius (public)	13	17
Respondent is in Cardiology (i.e. not orthopedics)	51.6	50.3
Respondent a physician (i.e. not a manager)	21.1	40.9
Respondent's tenure in the post (years)	3.50	3.79
Respondent's tenure in the trust (years)	10.28	8.56
Interview duration (minutes)	59.27	13.38
Number of patient-cases (per quarter)	15,513	8,207
Total employment	3,651.04	2,016.85
Number of sites	2.65	2.01
Medical Employees (Full-Time equivalent)	387.73	233.06

Notes: These are means and standard deviations for the sample of publicly owned acute hospital observations (NHS). There are usually 161 observations although exact number varies due to missing values.

Table 2: Hospital Performance and management practices

Dependent Variable:	(1) Mortality rate from emergency AMI	(2) Mortality rate from all emergency surgery	(3) Mortality rate from selected (high risk) emergency surgery	(4) Total waiting list	(5) Proportion in waiting list "at risk" of breaching national target	(6) MRSA infection rate	(7) Average intention of staff to leave in next 12 months	(8) Health Care Commission (HCC) overall rating	(9) "Pseudo" HCC rating
Management Practices Score (average over 16 Questions)	-0.063** (0.024)	-0.014** (0.006)	-0.128* (0.067)	-0.135*** (0.035)	-0.221*** (0.076)	-0.106 (0.092)	-0.214** (0.109)	0.421*** (0.093)	0.388*** (0.097)
Observations	140	160	153	160	160	160	160	161	161
Management Practices Score (average over 18 Questions)	-0.060** (0.025)	-0.015** (0.006)	-0.111 (0.068)	-0.129*** (0.036)	-0.230*** (0.076)	-0.118 (0.093)	-0.228** (0.105)	0.375*** (0.089)	0.421*** (0.106)
Observations	140	160	153	160	160	160	160	161	161

Notes: All dependent variables are standardized to be mean zero and standard deviation 1. The dependent variables in columns (1) through (7) are generally considered to be "bad" whereas those in (8) and (9) are "good" – see text for more details. Management scores are also standardized across the questions in Appendix A. These are OLS regressions with standard errors that are clustered at a hospital level (the unit of observation is a management interview with a service line in cardiology or orthopaedics across 100 public acute hospitals). *** significant at 1% level; ** significance at 5%, * for significance at 10%. All columns include "general controls" whether the respondent was a manager or clinician, speciality dummy, 10 regional dummies and the number of total admissions at the hospital level. Controls for case mix are also included, but vary across columns (see text for discussion). All columns also include "noise controls" comprising interviewer dummies, duration of the interview, number of questions not answered and tenure of the interviewee. The observations are weighted by the inverse of the number of interviews with the same hospital. Column (8) is average of HCC's rating on resource use and quality of service. Column (9) is our self-constructed HCC rating based on several indicators.

Table 3: Management Practice Regressions: Comparing across sectors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample	UK only	UK only	UK only	UK only	UK only	UK only	EU countries	EU countries	EU countries
Dependent variable	Management	management	management	management	management	management	management	management	management
(Type)	All	All	All	Monitoring	Targets	People	All	All	All
Manufacturing	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
Any Hospital	-0.529***	-0.814***	-0.972***	-0.470***	-0.904***	-0.772***	-0.561***	0.119	-0.546**
	(0.169)	(0.157)	(0.213)	(0.162)	(0.175)	(0.150)	(0.168)	(0.206)	(0.219)
Private organization								0.821***	0.307*
								(0.150)	(0.159)
Private hospital		1.617***		0.874***	1.213***	1.871***			1.409***
		(0.189)		(0.168)	(0.186)	(0.213)			(0.252)
Size			0.064						
(employees)			(0.067)						
Observations	833	833	753	833	833	833	1,993	1,993	1,993
NHS hospitals	161	161	146	161	161	161	161	161	161
Private hospitals	21	21	0	21	21	21	21	21	21
Manufacturing	651	651	607	651	651	651	1,811	1,811	1,811

Notes: ***represents significant at the 1% level; **significance at 5%, *significance at 10%. Dependent variable is standardized management score. Management "Type" is whether we average over 16 questions (excluding 2 questions on lean operations) or look at a sub-category (see Appendix A): Monitoring: question 4-6, Targets: question 8-12, People management: question 7 and 13-18. These are coefficients from OLS regressions with robust standard errors that are clustered at a hospital level (the unit of observations is a service line in cardiology or orthopaedics). Any hospital includes private and public hospitals, private organization includes private hospitals. EU includes manufacturing firms in France, Germany, Italy, Sweden and the UK. All regressions include multinational controls (dummies equal to one if the firm is a domestic or foreign multinational) and Noise controls (interviewer dummies, the duration of the interview and the tenure of the interviewee). The observations are weighted by the inverse of the number of interviews with the same hospital. See text for more discuss

Table 4: Management Practice Regressions, UK public hospitals

Dependent variable	(1) Management	(2) Management	(3) Management	(4) Management
Foundation hospital	0.533***			0.633***
	(0.192)			(0.180)
Proportion of managers		0.820**		0.926**
with a clinical degree		(0.355)		(0.343)
Size (number of patient-			0.541	1.689*
Cases/10000)			(0.972)	(0.919)
Observations	161	161	161	161

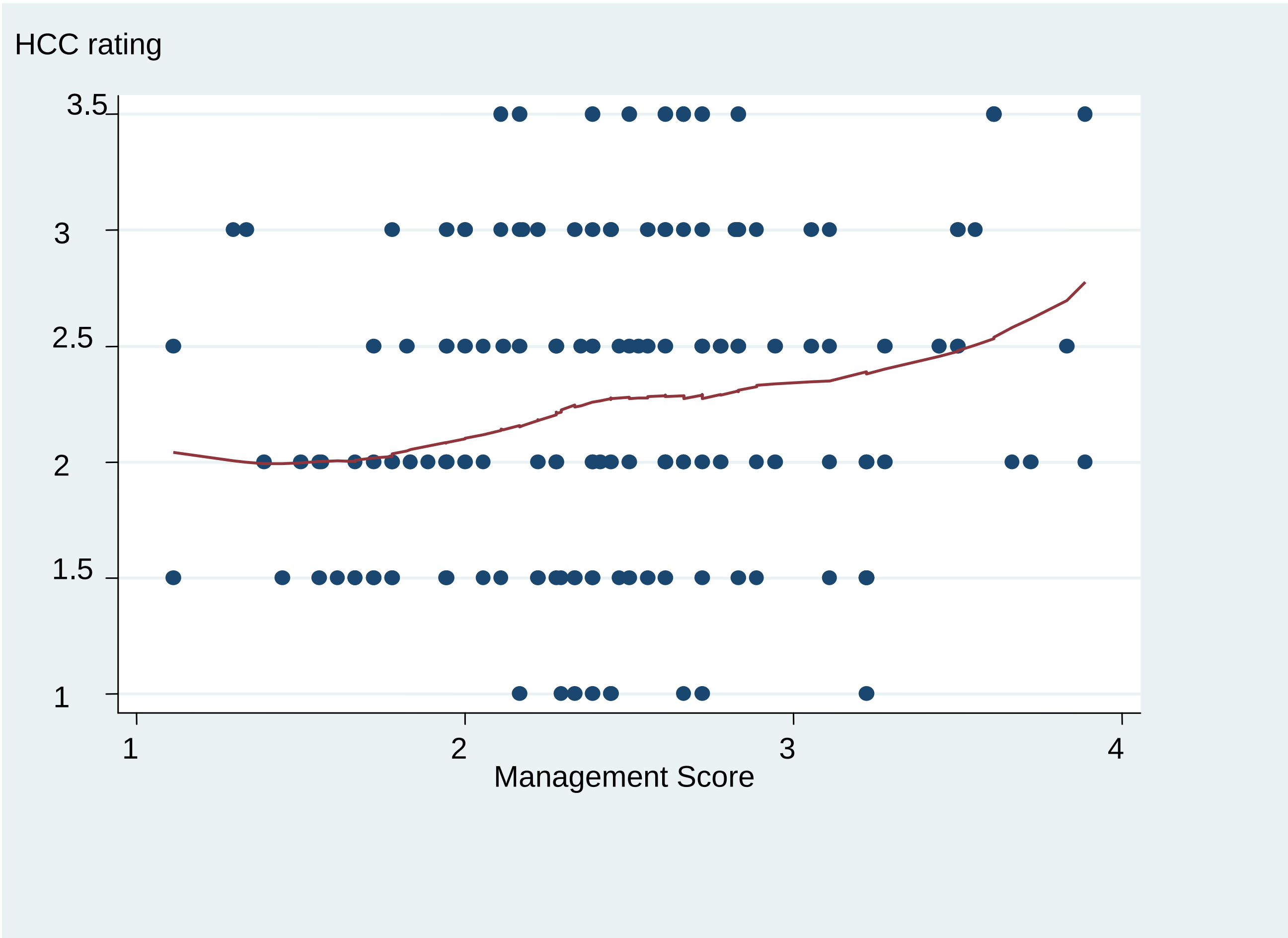
Notes: These are OLS regressions with robust standard errors that are clustered at a hospital level (the unit of observation is a management interview with a service line in cardiology or orthopaedics. *** significant at 1% level; ** significance at 5%, * for significance at 10%. Management is standardized over 16 questions. All columns include "general controls" whether the respondent was a manager or clinician, speciality dummy, 10 regional dummies and the number of total admissions at the hospital level. All columns also include "noise controls" comprising interviewer dummies, duration of the interview, number of questions not answered and tenure of the interviewee. The observations are weighted by the inverse of the number of interviews with the same hospital. See text for more details.

Table 5: Management Practice Regressions, Competition variables

	(1)	(2)	(3)
Dependent variable:	Standardized Average HCC rating	Standardized Average HCC rating	Standardized management score
Competition	0.043**	0.035*	0.016*
	(0.020)	(0.020)	(0.009)
Management score		0.493***	
		(0.110)	
Sample	Public	Public	Public
Observations	159	159	159

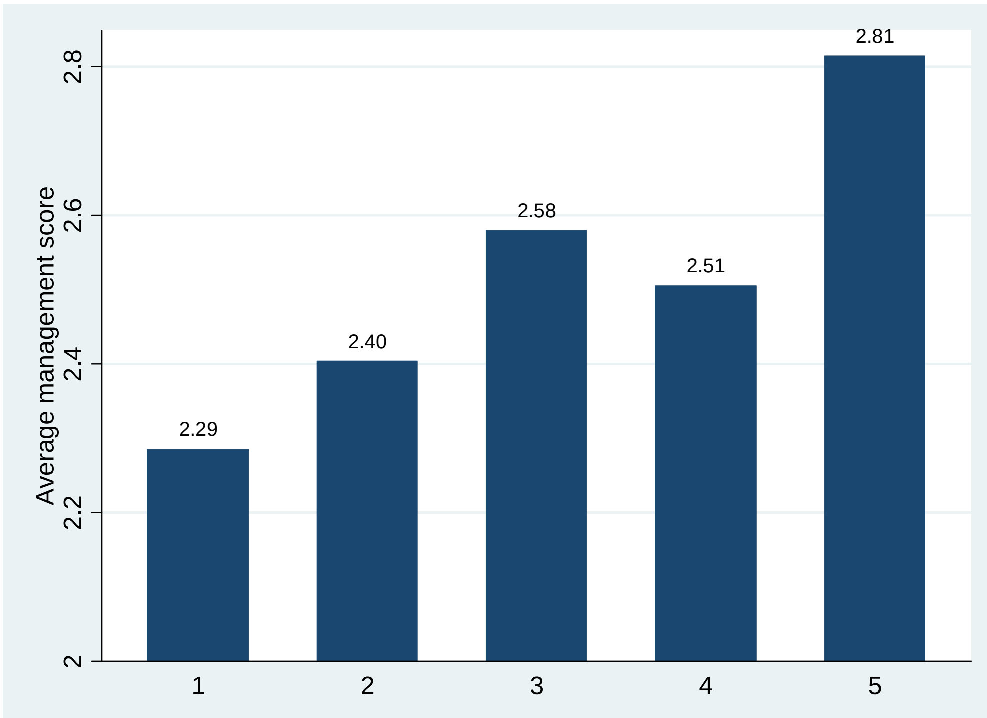
Notes: Dependent variable is the (standardized) average HCC rating (column (1) and (2)) and the (standardized) management score for 16 questions (column (3)). Competition is measured as the number of hospitals in a 30km radius around the hospital. These are OLS regressions with robust standard errors that are clustered at a hospital level (the unit of observations is a service line in cardiology or orthopaedics, so we have up to four observations in each hospital). (***) represents coefficients that are significant at 1%; (**) stands for significance at 5%, (*) for significance at 10%. We include casemix controls (age-gender proportions admitted and area mortality rate), a speciality dummy, number of patient cases, proportion of managers with clinical degree and 10 regional dummies and the noise controls (interviewer dummies, dummy for whether the respondent was a manager or clinician, duration of the interview, the number of management questions not answered and the tenure of the interviewee). The observations are weighted by the inverse of the number of interviews with the same hospital. See text for more discussion.

Figure 1: Average HCC score and management score



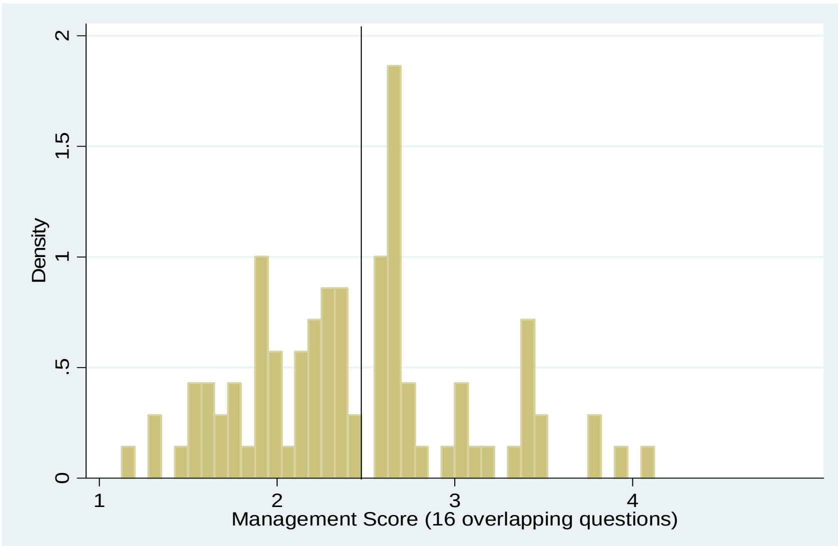
Notes: Each point represents a survey response. Vertical axis shows the average HCC score on "resource use" and "quality of service" in 2005/2006 (original data range is 1 to 4). Horizontal axis is the average management score over the 18 questions. The line is the local linear regression line.

Figure 2: Management Score by quintiles of average HCC rating



Notes: We divide the HCC average score into quintiles from lowest score (first) to highest score (fifth). We show the average management score (over all 18 questions) in each of the quintiles. The better performing hospitals have higher management scores.

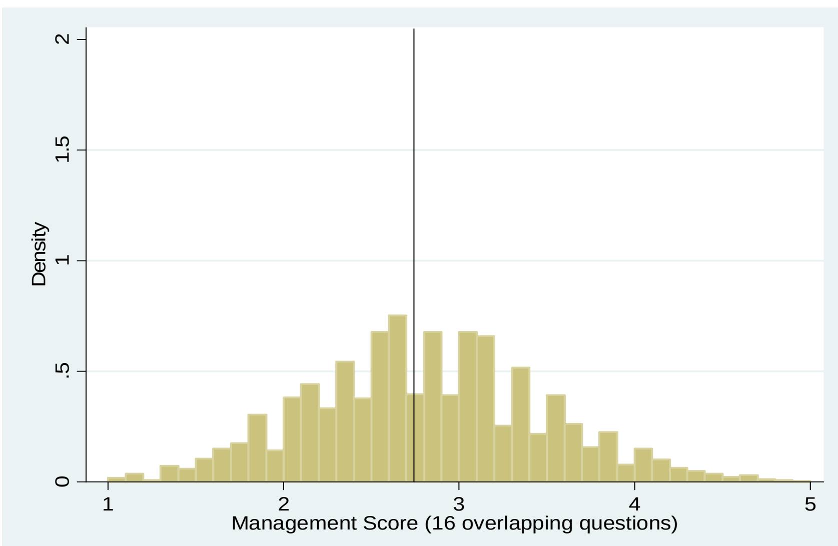
Figure 3: Comparison of Management Scores in Hospitals and Manufacturing Firms



Panel A

Management Scores

in Hospitals



Panel B:

Management Scores

In manufacturing firms

Notes: These are the distributions of the management score for hospitals and manufacturing firms. Only establishments who have between 50 and 5000 employees and who are domestically owned (i.e. multinationals were dropped from the manufacturing sample) were used here. Also observations with a low reliability score (below 3) were dropped. The vertical line represents the average management score in each sample. Only the 16 questions for which manufacturing and healthcare are comparable were used.

APPENDIX A: MANAGEMENT PRACTICE INTERVIEW GUIDE FOR THE HEALTHCARE SECTOR

Any score from 1 to 5 can be given, but the scoring guide and examples are only provided for scores of 1, 3 and 5. Multiple questions are used for each dimension to improve scoring accuracy.

(1) Lay out of patient flow

Tests how well the patient pathway is configured at the infrastructure level and whether staff pro-actively improve their own work-place organisation

a) Can you briefly describe the patient journey or flow for a typical episode?
b) How closely located are wards, theatres, diagnostics centres and consumables?
c) Has the patient flow and the layout of the hospital changed in recent years? How frequently do these changes occur and what are they driven by?

	Score 1	Score 3	Score 5
Scoring grid:	Lay out of hospital and organisation of workplace is not conducive to patient flow, e.g., ward is on different level from theatre, or consumables are often not available in the right place at the right time	Lay out of hospital has been thought-through and optimised as far as possible; work place organisation is not regularly challenged/changed (or vice versa)	Hospital layout has been configured to optimize patient flow; workplace organization is challenged regularly and changed whenever needed

(2) Rationale for introducing standardisation/ pathway management

Test the motivation and impetus behind changes to operations and what change story was communicated

a) Can you take me through the rationale for making operational improvements to the management of patient pathway? Can you describe a recent example?
b) What factors led to the adoption of these practices?
c) Who typically drives these changes?

	Score 1	Score 3	Score 5
Scoring grid:	Changes were imposed top down or because other departments were making (similar) changes, rationale was not communicated or understood	Changes were made because of financial pressure and the need to save money or as a (short-term) measure to achieve government targets	Changes were made to improve overall performance, both clinical and financial, with buy-in from all affected staff groups. The changes were communicated in a coherent ‘change story’

(3) Continuous improvement

Tests process for and attitudes to continuous improvement and whether things learned are captured/documented

a) How do problems typically get exposed and fixed?
b) Talk me through the process for a recent problem that you faced
c) How do the different staff groups get involved in this process? Can you give examples?

	Score 1	Score 3	Score 5
Scoring grid:	No, process improvements are made when problems occur, or only involve one staff group	Improvements are made irregular meetings involving all staff groups, to improve performance in their area of work (e.g., ward or theatre)	Exposing problems in a structured way is integral to individuals' responsibilities and resolution involves all staff groups, along the entire patient pathway as a part of regular business processes rather than by extraordinary effort/teams

(4) Performance tracking

Tests whether performance is tracked using meaningful metrics and with appropriate regularity

a) What kind of performance indicators would you use for performance tracking?
b) How frequently are these measured? Who gets to see these data?
c) If I were to walk through your hospital wards and theatres, could I tell how you were doing against your performance goals?

	Score 1	Score 3	Score 5
Scoring grid:	Measures tracked do not indicate directly if overall objectives are being met, e.g., only government targets tracked. Tracking is an ad-hoc process (certain processes aren't tracked at all).	Most important performance indicators are tracked formally; tracking is overseen by senior staff.	Performance is continuously tracked and communicated against most critical measures, both formally and informally, to all staff using a range of visual management tools

(5) Performance review

Tests whether performance is reviewed with appropriate frequency and communicated with staff

a) How do you review your KPI's?
b) Tell me about a recent meeting
c) Who is involved in these meetings? Who gets to see the results of this review?
d) What is the follow-up plan?

	Score 1	Score 3	Score 5
Scoring grid:	Performance is reviewed infrequently or in an un-meaningful way e.g. only success or failure is noted	Performance is reviewed periodically with both successes and failures identified. Results are communicated to senior staff. No clear follow up plan is adopted.	Performance is continually reviewed, based on the indicators tracked. All aspects are followed up to ensure continuous improvement. Results are communicated to all staff.

(6) Performance dialogue

Tests the **quality** of review conversations

a) How are these meetings structured?
b) During these meetings do you find that you generally have enough data?
c) What type of feedback occurs in these meetings?

	Score 1	Score 3	Score 5
Scoring grid:	The right information for a constructive discussion is often not present or the quality is too low; conversations focus overly on data that is not meaningful. Clear agenda is not known and purpose is not explicitly. Next steps are not clearly defined	Review conversations are held with the appropriate data present. Objectives of meetings are clear to all participating and a clear agenda is present. Conversations do not, drive to the root causes of the problems, next steps are not well defined	Regular review/performance conversations focus on problem solving and addressing root causes. Purpose, agenda and follow-up steps are clear to all. Meetings are an opportunity for constructive feedback and coaching

(7) Consequence management

Tests whether differing levels of (personal) performance lead to different consequences (good or bad)

a) Let's say you've agreed to a follow up plan at one of your meetings, what would happen if the plan weren't enacted?
b) How long is it between when a problem is identified to when it is solved? Can you give me a recent example?
c) How do you deal with repeated failures in a specific sub-specialty or cost area?

	Score 1	Score 3	Score 5
Scoring grid:	Failure to achieve agreed objectives does not carry any consequences	Failure to achieve agreed results is tolerated for a period before action is taken	A failure to achieve agreed targets drives retraining in identified areas of weakness or moving individuals to where their skills are appropriate

(8) Target balance

Test whether targets cover a sufficiently broad set of metrics

a) What types of targets are set for the hospital? What are the goals for your specialty?
b) Tell me about goals that are not set externally (e.g. by the government, regulators).

Score 1	Score 3	Score 5

Scoring grid:	Goals focussed only on government targets and achieving the budget	Goals are balanced set of targets (including quality, waiting times, operational efficiency, and financial balance). Goals form part of the appraisal for senior staff only or do not extend to all staff groups. Real interdependency is not well understood	Goals are a balanced set of targets covering all four dimensions (see left). Interplay of all four dimensions is understood by senior and junior staff (clinicians as well as nurses and managers)

(9) Target inter-connection

Tests whether targets are tied to hospital/Trust objectives and how well they cascade down the organisation

a) What is the motivation behind your goals?
b) How are these goals cascaded down to the different staff groups or to individual staff members?
c) How are your targets linked to hospital performance and its goals?

	Score 1	Score 3	Score 5
Scoring grid:	Goals do not cascade down the organisation	Goals do cascade, but only to some staff groups, e.g., nurses only	Goals increase in specificity as they cascade, ultimately defining individual expectations, for all staff groups

(10) Time horizon of targets

Tests whether hospital/Trust has a ‘3 horizons’ approach to planning and targets

a) What kind of time scale are you looking at with your targets?
b) Which goals receive the most emphasis?
c) Are the long term and short term goals set independently?
d) Could you meet all your short-run goals but miss your long-run goals?

	Score 1	Score 3	Score 5
Scoring grid:	Top staff’s main focus is on short term targets	There are short and long term goals for all levels of the organisation. As they are set independently, they are not necessarily linked to each other	Long term goals are translated into specific short term targets so that short term targets become a ‘staircase’ to reach long term goals

(11) Target stretch

Tests whether targets are appropriately difficult to achieve

a) How tough are your targets? Do you feel pushed by them?
b) On average, how often would you say that you meet your targets?
c) Do you feel that on targets all specialties, departments or staff groups receive the same degree of difficulty? Do some groups get easy targets?
d) How are the targets set? Who is involved?

	Score 1	Score 3	Score 5
Scoring grid:	Goals are either too easy or impossible to achieve, at least in part because they are set with little clinician involvement, e.g., simply off historical performance	In most areas, senior staff push for aggressive goals based, e.g., on external benchmarks, but with little buy-in from clinical staff. There are a few sacred cows that are not held to the same standard	Goals are genuinely demanding for all parts of the organisation and developed in consultation with senior staff, e.g., to adjust external benchmarks appropriately

(12) Clarity and comparability of targets

Tests how easily understandable performance measures are and whether performance is openly communicated

a) If I asked your staff directly about individual targets, what would they tell me?
b) Does anyone complain that the targets are too complex?
c) How do people know about their own performance compared to other people's performance?

	Score 1	Score 3	Score 5
Scoring grid:	Performance measures are complex and not clearly understood, or only relate to government targets. Individual performance is not made public	Performance measures are well defined and communicated; performance is public at all levels but comparisons are discouraged	Performance measures are well defined, strongly communicated and reinforced at all reviews; performance and rankings are made public to induce competition

(13) Managing talent

Tests what emphasis is put on talent management

a) How do senior staff show that attracting and developing talent is a top priority?
b) Do senior managers, clinicians or nurses get any rewards for bringing in and keeping talented people in the hospital?

	Score 1	Score 3	Score 5
Scoring grid:	Senior staff do not communicate that attracting, retaining and developing talent throughout the organisation is a top priority	Senior management believe and communicate that having top talent throughout the organisation is key to good performance	Senior staff are evaluated and held accountable on the strength of the talent pool they actively build

(14) Rewarding high performers

Tests whether good performance is rewarded proportionately

a) How does your appraisal system work? Tell me about your most recent round.
b) Are there any non-financial or financial (bonuses) rewards for the best performers across all staff groups?
c) How does the bonus system work?
d) How does your reward system compare to that at other comparable hospitals?

	Score 1	Score 3	Score 5
Scoring grid:	People are rewarded equally irrespective of performance level	There is an evaluation system for the awarding of performance related rewards that are non-financial (beyond progression through nursing grades or clinical excellence awards for doctors) at the individual level (but rewards are always or never achieved)	There is an evaluation system for the awarding of performance related rewards, including personal financial rewards

(15) Removing poor performers

Tests whether hospital is able to deal with underperformers

a) If you had a clinician or a nurse who could not do his job, what would you do? Could you give me a recent example?
b) How long would underperformance be tolerated?
c) Do you find staff members who lead a sort of charmed life? Do some individuals always just manage to avoid being fixed/fired?

	Score 1	Score 3	Score 5
Scoring grid:	Poor performers are rarely removed from their positions	Suspected poor performers stay in a position for a few years before action is taken	We move poor performers out of the hospital/department or to less critical roles as soon as a weakness is identified

(16) Promoting high performers

Tests whether promotion is performance based

a) Tell me about your promotion system?
b) What about poor performers? What happens with them? Are there any examples you can think of?
c) How would you identify and develop your star performers?
d) Are better performers likely to promote fasters or are promotions given on the basis of tenure/seniority?

	Score 1	Score 3	Score 5
Scoring grid:	People are promoted primarily on the basis of tenure	People are promoted upon the basis of performance (across more than one dimension, e.g., isn’t related only to research or clinical excellence)	We actively identify, develop and promote our top performers

(17) Attracting talent

Tests how strong the employee value proposition is

a) What makes it distinctive to work at your hospital, as opposed to your other similar hospitals?
b) If I were a top nurse or clinician and you wanted to persuade me to work at your hospital, how would you do this?
c) What don't people like about working at your hospital?

	Score 1	Score 3	Score 5
Scoring grid:	Our competitors offer stronger reasons for talented people to join their hospitals	Our value proposition to those joining our department is comparable to those offered by others hospitals	We provide a unique value proposition to encourage talented people join our department above our competitors

(18) Retaining talent

Tests whether hospital/Trust will go out of its way to keep its top talent

a) If you had a top performing manager, nurse or clinician that wanted to leave, what would the hospital do?
b) Could you give me an example of a star performer being persuaded to stay after wanting to leave?
c) Could you give me an example of a star performer who left the hospital without anyone trying to keep them?

	Score 1	Score 3	Score 5
Scoring grid:	We do little to try and keep our top talent	We usually work hard to keep our top talent	We do whatever it takes to retain our top talent across all three staff groups

APPENDIX B: Data

Sample

The main sampling frame was all acute public sector hospitals (NHS "trusts"25) in England. There were 174 such units in 2006, but we dropped hospitals without orthopaedics or cardiology departments (e.g. specialist eye hospitals) so this left us with a sample of 164 possible hospital trusts. We obtained 161 usable responses from 100 hospital trusts which represented 61% of the frame, so we essentially have the population. We sought responses from up to four senior employees in each hospital: a manager and a clinician from two service lines (cardiology and orthopaedics). Table 1 shows that we are split evenly between the specialities (52% cardiology and 48% orthopaedics), but also that it was harder to obtain interviews with the physicians than managers (80% of the respondents were managers). Table B3 shows the breakdown of the number of interviews by hospital: we only obtained one interview for 53 of the trusts.

We examined evidence for selection bias by estimating probit models of whether a trust responded on the observable characteristics. These characteristics are draw from Department of Health datasets (such as Hospital Episode Statistics). Table B4 contains the results of this exercise. There is no significant correlation between sample response and any of the performance measures or the covariates which suggests that there was little systematic response bias.

It is more difficult to carry out a similar exercise for the private hospitals as there is less information on the non-responding hospitals (public hospitals are required to lodge a large amount of data with the Department of Health whereas private hospitals are not). We were, however, able to obtain some data from Laing and Buisson 215 private sector hospitals where some form of orthopaedic or cardiological services were available. This would make our sample only about 10% of the total. The dataset only has basic information. We examined whether there was any correlation with the number of beds (a proxy for size). This variable was insignificant as was the number of day places.

In the regressions all interviews with a very short duration (less than 25 minutes) or many unanswered questions (at least 3) are excluded completely as the information obtained is not reliable.

We weight regressions by the inverse of the number of interviews so that hospitals with multiple responses are weighted less (we also cluster standard errors at the hospital level.

Construction the Pseudo HCC Rating

In column (9) of Table 2 we reported our best effort to reconstruct the HCC's rating. Although the exact method of creating the HCC ratings is not publicly known the

[25] A trust can consist of more than one site (as a firm can consist of more than one plant). The median number of sites was 2 with a range from 1 to 10.

Appendix of the HCC's “Annual Health Check 2006/2007” brochure mentions seven “domains” in which the hospitals need to achieve certain standards in order to achieve a high score.

These domains are: safety, clinical and cost effectiveness, governance, patient focus, accessible and responsive care, public health, and care environment and amenities. From the datasets described above we choose eight variables which capture the requirements of these different domains[26]. Infection rates and re-admission risk are chosen to represent the “safety” aspect; operational margin and income per medical FTE capture the financial side; patient satisfaction covers the “patient focus” domain. Waiting times and average length of stay fall into the category “accessible and responsive care” and information on job satisfaction from the NHS staff survey is used to represent the “care environment and amenities” domain.

[26] The only categories which are not covered are governance and public health. Governance is directly related to the management score and therefore should not be included as it is already part of the dependent variable. There is furthermore no information in the data that corresponds to the public health category.

Table B1: Data Sources for hospital performance data

Variable	Notes	Source
Mortality within 28 days of emergency admission for AMI (in hospital and out of hospital)	• During financial quarter • Defined according to NHS mortality rate Performance indicators (PIs) for 2001/02	ONS death records linked with Hospital Episode Statistics (HES), The NHS Information Centre for health and social care.
Mortality within 30 days of surgery for selected emergency procedures. - All specialties - General surgery only	• During financial quarter • Defined according to NHS mortality rate PIs for 2001/02	ONS death records linked with Hospital Episode Statistics (HES), The NHS Information Centre for health and social care.
MRSA rates	• During financial quarter • 2001/02 (q1) to 2004/05 (q3)	Health Protection Agency: Quarterly reporting results for clostridium difficile infections and MRSA bacteraemia
Waiting list size	• At start of quarter (as proxied by end of previous quarter)	Department of Health: Provider based waiting times/list statistics [a]
"Distance from target": % of specialty list at risk of breaching target if untreated by next census date	• At start of quarter (as proxied by end of previous quarter)	Department of Health: Provider based waiting times/list statistics [b]
Probability of leaving in next 12 months	Respondents are asked to rate chances of	NHS Staff Survey[c] (2006). 128,328 NHS staff responded and results are reported as average of scale by each trust

	leaving on a 1 to 5 scale.	
Healthcare Commission rating[c]	All trusts are scored on a scale of 1 to 4 on "resource use" and quality of "care"	Our main indicator averages over the two measures and standardizes. We also construct our own "pseudo" HCC rating from the underlying indicators (see Appendix B for full description)
Local authority all cause mortality rates	• Calendar year	Office of National Statistics 1995-2004
Casemix of admissions. These are specific to the conditions (AMI, surgery, etc.) considered. For the general performance indicators (like HCC rating) we use all patients admitted.	• Proportion of admitted patients in each sex-specific age band. 11 categories: 0-15, 16-45, 46-50, 51-55, 56-60, 61-65, 66-70, 71-75, 76-80, 81-85, >85. So so up to 22 controls.	Hospital Episode Statistics (HES), The NHS Information Centre for health and social care.

Notes: Mortality indicators, the MRSA indictor, waiting times and waiting lists have been used by the UK healthcare regulator to assess the performance of hospitals in the NHS.

[a] http://www.performance.doh.gov.uk/nhsperformanceindicators/2002/trdca_t.doc.

[b] http://www.performance.doh.gov.uk/waitingtimes/index.htm

[c] http://www.healthcarecommission.org.uk/guidanceforhealthcarestaff/nhsstaff/nhsstaffandpatientsurveys/staffsurveys.cfm

Table B2: Hospitals compared with manufacturing

	Variable	Mean	Std. Dev.	Mean UK manufacturing	Difference to manufacturing	Difference significant (1% level)
Overall	Average Management Score	2.57	0.66	2.96	-0.39	*
Subcategories	Average Operations Score	2.83	0.95	not applicable		
	Average Monitoring Score	3.00	0.75	3.32	-0.32	*
	Average Targets Score	2.47	0.78	2.93	-0.45	*
	Average People Man. Score	2.35	0.70	2.86	-0.51	*
Operations	Layout of patient flow	2.76	1.18	not applicable		
	Rational for standardisation	2.88	1.24	not applicable		
	Continuous improvement	2.90	1.12	3.13	-0.24	*
Monitoring	Performance tracking	2.97	0.94	3.38	-0.42	*
	Performance review	3.26	0.91	3.36	-0.10	
	Performance dialogue	2.77	0.95	3.21	-0.44	*
Targets	Target balance	2.33	1.26	2.94	-0.60	*
	Target inter-connection	3.01	1.15	3.01	0.00	
	Time horizon of targets	2.20	1.29	3.08	-0.89	*
	Target stretch	2.61	1.02	3.01	-0.39	*
	Clarity and comparability of targets	2.21	0.88	2.60	-0.39	*
People management	Consequence management	3.03	1.09	3.19	-0.15	
	Managing talent	1.71	1.01	2.43	-0.72	*
	Rewarding high performers	2.01	0.97	2.62	-0.61	*
	Removing poor performers	2.56	1.08	3.13	-0.57	*
	Promoting high performers	2.49	0.98	3.04	-0.55	*
	Attracting talent	2.85	0.99	3.08	-0.23	*
	Retaining talent	1.83	1.05	2.51	-0.69	*

Notes: These are tests are the difference between publicly owned NHS hospitals with manufacturing firms in the UK.

Table B3: Hospitals Interviewed

Number of Interviews per NHS hospital:

interviews	hospitals	Observations
1	53	53
2	34	68
3	12	36
4	1	4
Total	100	161

Notes: The unit of observation is an interview with either a manager or a clinician based at cardiology or orthopedics. Up to four interviews per hospital are possible. There were 161 interviews in 100 hospital trusts.

Table B4: Sample Selection for public hospitals?

Variable	Marginal effect (standard error)	Observations
Mortality rate from AMI	0.129(0.161)	133
Mortality rates from general surgery	0.239(0.334)	163
Mortality rates from selected high risk surgery	0.017(0.049)	152
Total Waiting List	0.025(0.045)	163
Proportion on waiting list "at risk"	0.026(0.060)	163
MRSA Infection rate	-0.025(0.041)	163
Health Care Commission overall rating	-0.014(0.056)	163
HCC Rating over sub-set of indicators	0.067(0.090)	164
Area Standardized Mortality Rate	0.0003(0.0003)	162
Number of patient cases/10000	0.018(0.470)	163
Number of employees/10000	-0.039(0.200)	164
Proportion doctors	-2.463(2.177)	154
Foundation Trust	0.091(0.082)	164

Notes: These are the results from separate probit ML regression of whether a public hospital had any response to the survey on the relevant variable (e.g. mortality rates in the first row). There is a population of 164 potential acute hospitals in England and we had 100 hospitals with at least one respondent.