The University of Manchester

The University of Manchester Research

Combining regression trees and panel regression for exploring and testing the impact of complementary management practices on short-notice elective operation cancellation rates

DOI:
10.1080/20476965.2019.1596338

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):
Salehnejad, R., Ali, M., & Proudlove, N. (2019). Combining regression trees and panel regression for exploring and testing the impact of complementary management practices on short-notice elective operation cancellation rates. *Health Systems*, *9*(4), 326-344. https://doi.org/10.1080/20476965.2019.1596338

Published in:
Health Systems

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

OPEN ACCESS

Download date:08. Feb. 2025

Combining regression trees and panel regression for exploring and testing the impact of complementary management practices on short–notice elective operation cancellation rates

December 2018

Abstract

Variation in the performance of providers across healthcare systems is pervasive. It is recognised as both a major concern and an opportunity for learning and improvement. Variation between providers is broadly considered to be due to management practices and contextual factors such as catchment-area demographics. However, there is little understanding of the ways in which these impact on performance and how they can be measured. We use recent developments in both regression trees and panel regression techniques to explore and then statistically test complementary alignments of management practices whilst taking into account contextual factors. We apply this to five years of NHS hospital trust data, examining performance on short–notice cancellation rates. We find that different alignments of management practices give rise to quite different short–notice cancellation rates between trusts, with some being substantially lower. Our research offers a data–driven approach for identifying optimal clusters of management practices.

Keywords : Cancelled Elective Operations, Complementarity, Hospital Performance, Management Practices, Regression Trees, Panel Regression

Introduction

Variation in performance of providers across healthcare systems is pervasive and recognised as a major concern and opportunity for improvement (Carter, 2016; Castelli et al., 2015; Chandra et al., 2016). This paper studies the variation in rates of short–notice cancellations of elective (i.e. non-emergency) operations across English NHS acute hospital trusts and explores potential drivers. A wealth of literature suggests the critical importance of management practices and styles, organisational culture and other contextual factors in driving performance in healthcare organisations (Kaplan et al., 2010). However, there is little understanding of the ways in which these impact on performance and how they can be measured (Harvey et al., 2015).

We put together a panel dataset on short–notice cancellation rates and a rich set of management practices in NHS hospital trusts, spanning five financial years. The sample also includes relevant control variables including trusts effects. The data set is 'wide' in that it includes very many variables. The paper aims to explore systematically–interacting structures (complementarities) among the variables, and threshold levels in them, that drive short–notice cancellation rates. We use exploratory techniques as a first stage to guide our regression analysis of complementarities.

We employ a regression tree approach as an exploratory tool to identify relevant clusters of management practices in a data-driven manner. Tree-based techniques have previously been used to group patient types for predicting radiology procedure times (Huang and Marcak, 2013), splitting inpatients into homogeneous groups based on length of stay (Harper and Shahani, 2002), identifying determinants of infection (Glowacka et al., 2009) and classifying decision makers' strategies (Robinson et al., 2005). In this paper we use a recent development in the machine learning literature, unbiased regression trees, to identify potentially important clusters of management practices. We also replicate the analyses using clustering techniques, finding very similar patterns.

We then use correlated random effects panel regression to test for statistical significance of the clusters, using dummy variables to indicate cluster membership and interactions with other management practices and variables relating to other trust characteristics such as workforce skill-mix. This novel combination of two recently-developed techniques provides a powerful method for systematically searching for then testing optimal combinations of variables, currently lacking in the management science literature.

We find rich complementarities amongst management practices and strong evidence on possible impacts of clusters or, more precisely, of alignments of complementary practices on short–notice cancellation rates, where an *alignment* refers to a cluster of complementary management practices that are set at the 'right' level. The findings strengthen our previous research on managerial and organisational determinants of trust performance (Ali et al., 2018a,b).

Our paper contributes to a growing literature that traces, at a fundamental level, variations in hospital performance across healthcare systems to differences in management practices (Bloom et al., 2013; Lega et al., 2013). Our key insight is that the unit of analysis in management studies is a *complementary cluster* of management practices rather that individual practices. We document substantial evidence using panel regression that clusters of management practices reduce or increase performance. The analysis has policy implications. Any effort at reducing short–term cancellation rates should eventually involve putting in place a *cluster* of *complementary* practices that optimises performance. Changing clusters of management practices would be a major challenge, which perhaps contributes to the persistence in the performance of trusts that we observe.

This paper is organised as follows. We first review the evidence on causes and consequences of short–notice cancellations, then outline a conceptual framework and state our research hypotheses. We then describes the data and define relevant measures. Next we introduce the unbiased panel regression tree technique and use it to explore complementarities, then test these with panel regressions. After interpreting the results we finish with a discussion and conclusions.

Short–notice cancellations of elective operations

A particular issue in the treatment of elective patients is the cancellation of some procedures at short–notice, often once a patient has already been admitted to hospital and in some cases on the day that the procedure was scheduled to take place. Other terms used are last–minute or On The Day of Operation cancellations. Short–notice cancellation rates exceeding 10% of all scheduled cases have been reported for public and private hospitals in the United States (Argo et al., 2009; Hand et al., 1990; Lacqua and Evans, 1994) and in Europe (Mangan et al., 1992; Sanjay et al., 2007; Seim et al., 2009). In Germany, cancellation rates have been found to vary from 0.8% to 17.9% between hospitals (Schuster et al., 2011). A number of studies have shown that as much as 60-80% of short–notice cancellations are avoidable (e.g. Rymaruk (2011)).

Estimating the full costs arising from short–notice cancellations is very difficult. In the English NHS prior to 2012, short–notice cancellations received a specific Healthcare Resource Group code, with an associated fixed reimbursement tariff (Cookson et al., 2017), allowing an estimation of direct costs of nearly £70 million for 2007 (McIntosh et al., 2012). Direct financial costs underestimate the full effects of short–term cancellations. Cancellations have a considerable impact on the experience of patients and their families (Dimitriadis et al., 2013; Hovlid et al., 2013; Ivarsson et al., 2004; Schofield et al., 2005; Tait et al., 1997), may be clinically harmful (Argo et al., 2009; Ivarsson et al., 2002) and can reduce productivity of operating theatres and staff (Dimitriadis et al., 2013; Schofield et al., 2005). Every cancellation is a lost opportunity to provide efficient and high-quality care.

Although the overall financial and emotional impacts of short–notice cancellations could be substantial, there is very little known beyond several single–trust studies about their extent and causes of (McIntosh et al., 2012). The literature has identified a multitude of reasons for short–notice cancellations of elective operations. The commonest reasons include patients being not fit for the operation (Dimitriadis et al., 2013; Haana et al., 2009), bed unavailability due to increased number of emergency admissions (Nasr et al., 2004; O'Malley et al., 2005), lack of theatre time (Pollard and Olson, 1999), the intervention no longer being necessary (Haana et al., 2009; Rai and Pandit, 2003) and patients failing to attend (Basson et al., 2006; Singh et al., 2005). Cookson et al. (2013) and McIntosh et al. (2012) identify age, gender, day of admission, socio–economic status and hospital characteristics as significant predictors of short–notice cancellations.

In the English NHS, short–notice cancellations are classified as being either for *Clinical Reasons* (to do with the patient's fitness and need for surgery), *Patient Reasons* (e.g. failure to attend) or *Hospital Reasons* (to do with availability of resources such as theatre time, diagnostics, post–operative intensive or critical care beds). Only *Hospital Reasons* data are published nationally, at hospital–trust level (a trust may run one or several acute hospitals). A study in an NHS trust in 2012 revealed an overall cancellation rate of 5%, with over half of these (3% of operations) being cancelled for *Hospital Reasons*, with leading causes being lack of beds (22%) and lack of theatre time (17%) (Dimitriadis et al., 2013). Another similar study produced an estimated overall cancellation rate of 13% with approximately a third (4.5% of all operations) cancelled for *Hospital Reasons* (Sanjay et al., 2007). A more recent study of a short window of data from many hospitals produced a very similar overall rate (Wong et al., 2018). The data also show strong persistence of high or low trust performance over many years (Proudlove et al., 2018b).

Research Hypotheses

What drives the dispersion in the rate of short–notice cancellations of elective surgery across NHS trusts? The literature above suggests a large range of causes from bed unavailability, through to failure of patients to attend or patients being unfit for the operation. However, these factors are simply *proximate* causes of cancellations. They could themselves be the effects of more fundamental causes. For example shortage of theatre time can be due to overambitious theatre lists because of surgeons underestimating theatre time required (Dimitriadis et al., 2013), because of inappropriate management of the waiting list (Schofield et al., 2005), shortage of pre–assessment capacity (Rai and Pandit, 2003), or lack of coordination between the various disciplines involved along the care pathways. There could be poor capacity planning for hospital beds and ineffective planning for emergency demand (Proudlove et al., 2003).

All these considerations suggest that, at a deeper level, cancellations may be at least partly due to ineffective management more widely (Dimitriadis et al., 2013). Not only does availability of resources matter, the management and organisation of these resources may also play an important part. It is inadequate to focus purely on immediate causes of cancellation and overlook more fundamental reasons. Any effort at reducing short–notice cancellations inevitably requires dealing with more fundamental factors relating to trusts' internal organisational design and management practices.

This paper thus distinguishes between *immediate* and *deep* drivers of last–minute cancellation rates, where by deep drivers we refer to managerial and organisational factors that affect trust performance. Management practices are strongly complementary in that "doing (more of) any one of them increases the return to doing (more of) the others" and appear as a cluster in an organisation (Milgrom and Roberts, 1995, p.181). Our conjecture, consequently, is that variation in clusters of complementary management practices drives variation in trust performance in general and in the specific case of cancellation rates.

Our general hypothesis is that highly–efficient trusts experience a lower cancellation rate because they manage their resources and coordinate their activities more effectively. The emphasis on management practices in driving trust efficiency is not new. A substantial body of literature documents evidence of their impact (Bloom et al., 2015, 2013). The literature, however, has so far mainly sought to identify factors that drive performance and has paid little attention to interactions among the management practices. Theory, however, suggests that management practices strongly complement each other; and their total impact is larger than the sum of the parts when utilised in conjunction with each other (Holmstrom and Milgrom, 1994; Ichniowski et al., 1997). Teamwork, for example, is more effective in organisations when it is complemented by other advanced management practices such as group incentives, training, effective communication and flexibility (Milgrom and Roberts, 1995). For example, in the absence of an effective group incentive system, teams may experience free–riding challenges and lower performance. Further, theory points to the importance of critical interactions between management/organisational practices and other inputs (human capital, skills and physical assets) in driving trust performance, not simply the mere presence or absence of the factors or their levels (Brynjolfsson and Milgrom, 2013). With this in mind, our basic research hypothesis is:

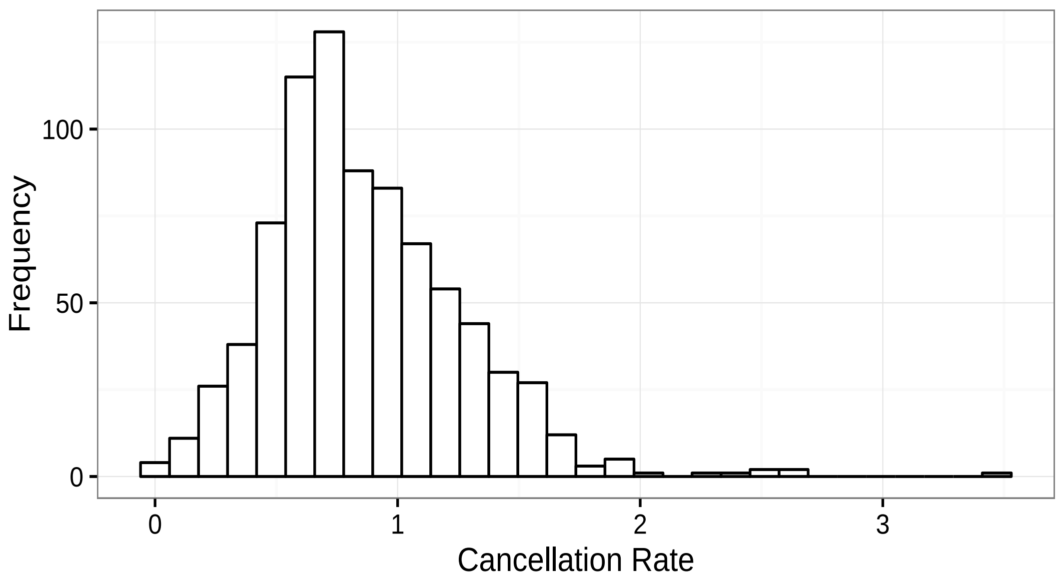
Research Hypothesis: Alignments of management practices matter in driving hospital organisation (here trust) efficiency. Specifically, alternative alignments of management practices give rise to different short–notice cancellation rates, with some leading to substantially lower rates.

Further, we refer to a complementary set of management practices that is adequate for ensuring performance as a *sufficient* complementary set. It is plausible to think that a *sufficient* set of complementary practices may already be in place but performance still falls below its 'desired' level. This suggests that not only the presence of a sufficient complementary cluster of management practices is required for performance, but the strength or quality of the practices must also reach certain levels for them to have an optimal impact on performance. Clusters of complementary management practices may be present but at weaker levels, and so may fail to have any impact at all. These intuitions suggest the impact of management practices on performance may be *non–linear* in two ways. When some members of a sufficient complementary set are absent, the incomplete cluster of practices may have no impact on performance at all. And, even when all members of a sufficient complementary set are present, the impact on performance may still be nil as some of the practices are far from their 'optimal' level.

Data & Measures

We put together a longitudinal (panel) dataset on 173 English NHS hospital trusts, spanning the five financial years 2008/09 – 2012/13. The minimum and maximum number of trusts per year in our sample are 161 and 169 respectively. This variation is due to some trusts being dissolved, merged, or simply, having incomplete data. Our dependent variable is *Hospital Reasons* cancellation rates, defined as the number of operations "cancelled at the last minute for non–clinical reasons by NHS providers" (NHS England, 2016, p.2) divided by the total trust elective activity. Data are collated and published nationally at trust level quarterly (cancellations, the numerator) and monthly (denominator). We use these to derive annual (April–March financial year) figures for each NHS trust. From here on we use the term cancellation rate to refer to this metric. Figure 1 shows the variation in these cancellation rates across hospital trusts.

Figure 1: Variation in Short–Notice Cancellation Rates across NHS Trusts (N=173 trusts, 2008/09 – 2012/13). The cancellation rate for trusts falling in the third quartile (1.09) is almost twice as high as the rate for trusts falling in the first quartile of the distribution (0.59).



Although the average cancellation rate (0.84%) may appear low, the extent of the variation is considerable, for example trusts in the worst-performing decile have cancellation rates about four times worse than the best decile (1.6% vs 0.4%). The observed dispersion is very much greater than randomness would suggest and, moreover, the dispersion is highly persistent over time. Regressing the measure of cancellation rate in our sample on its one–year lag yields an autoregressive coefficient 0.652 – a high– or low–performing trust one year is very likely to remain high– or low–performing in subsequent years (see Figure S1 in the online supplement for further evidence). While our data cover all hospital trusts, the analysis will exclude specialist trusts, e.g., cancer and children's trusts, as their structures and workloads differ greatly.

Alongside cancellations, we have several sets of potential 'explanatory' variables, described below and in Tables 1 and 2.

Management, Culture and Environment: A growing body of theory and empirical evidence points to the possible role of management practices in driving performance (Ahmad and Schroeder, 2003; Birdi et al., 2008; Bryson et al., 2017; Chuang et al., 2012; West et al., 2006). We draw on the annual NHS Staff Survey (NSS) to obtain data on measures of management practices. The NSS is a mandatory survey for all acute NHS trusts. It has run annually since 2003 and is the largest survey of staff opinion in the UK. It captures views of clinical and non–clinical NHS staff on a range of such issues related to work experiences and staff well–being (Department of Health, 2011). Of principal interest are four variables: the decentralisation of decision making (*Decisions*), whether senior managers act on staff feedback (*Feedback*), suggesting ideas for improving services (*Ideas*) and effective communication between senior management and staff (*Communication*). These four practices constitute some of the key elements of Appelbaum et al. (2000)'s high–performance work system and measure staff perceptions of the management environment. Responses on these survey questions are Likert–type variables measured on a five–point scale from strongly disagree to strongly agree. The scores for these practices are constructed by taking their Positive Response Rate (PRR) i.e. the percentage of respondents who agree or strongly agree.

While our analysis mainly uses individual practices and interactions between them, we also construct a composite management score from them using principal component analysis. Figure 2 shows the scatter plot of the mean cancellation rate for each trust versus the mean of its composite management score over the sample period. The data points are labelled by the three-digit alphanumeric NHS trust codes. The regression line suggests a material and significant relationship: higher management scores are associated with lower cancellation rates, and so there appears to be a potentially interesting effect of trust management.

In addition, we include several other indices from the NSS as follow: *Team Quality* assesses team effectiveness, reflecting whether team members often discuss this, communicate closely with each other and have a set of shared objectives. *Flexible* assesses the use of flexible working options, such as working flexi–time, part–time, annualised hours, and teams making their

Figure 2: Relationship between cancellation rates and mean management practices score for trusts. The regression equation for the first graph is $y = 1.20 - 0.007x$, where $R^2 = 0.01$. The fitted equation for the second graph is $y = 1.24 - 0.008x$, where $R^2 = 0.13$. The graphs also show 95% confidence intervals (grey shaded areas) for these regression fits.

own decisions about rotas and job sharing. *Intention to Leave* assesses the extent to which staff are considering leaving their organisation, and looking for a new job either within or outside of the NHS. *Job Satisfaction* reflects employees' satisfaction with the recognition they receive for good work, the support they receive from their manager, and the extent to which the employer values their work. *Job Design* assesses the extent to which staff are performing jobs that have clear goals, give them opportunities to participate in decision making and provide clear feedback on performance. *Errors* assesses the extent to which staff are aware of the procedures for reporting errors, are encouraged to report errors and feel secure in disclosing incidents. *Appraisal* assesses the percentage of staff that had well-structured appraisal reviews within the previous 12 months. *Supervisor* assesses the extent to which staff receive support, guidance and feedback and are involved in decision making.

We also include potential confounders as control variables:

Trust Characteristics: We include dummy variables for *Teaching Trust* and *Foundation Trust* status. Foundation Trusts are permitted greater financial and managerial autonomy from direct government control. They are allowed to keep surpluses, which they can use to increase staff salaries and invest in facilities or improved services for patients (Allen et al., 2010). Teaching Trusts might incur higher costs and treat more complex patients. Moreover, requirements for consultants to spend time training medical students might lengthen treatment processes (Castelli et al., 2015). We also include the number of *Hospital Beds* as a measure of trust size. Cookson et al. (2017) and McIntosh et al. (2012) include these structural characteristics in their study of NHS cancellation rates.

Workforce Skill–mix: We also include the proportions of medical, nursing and support staff in the workforce, where support staff include nursing assistant practitioners, healthcare assistants and support workers. A higher proportion of medical staff could be related to lower cancellation rates since greater concentration of doctors increases the supply of skills, enables greater specialisation and division of labour and more decision making time with patients.

Patient Characteristics: Our dataset also includes a number of covariates describing patient characteristics to prevent any heterogeneity bias. These variables are the percentages of total admissions categorised as emergencies, female and aged 60 plus. Higher older and emergency patient demand could increase variety and variability, with knock–ons to elective flows (McIntosh et al., 2012).

Geographical Characteristics: To control for regional factors, we add the Market Forces Factor (MFF) for each trust's location. The MFF captures unavoidable geographic differences in costs facing providers such as staff costs and capital (e.g., land and buildings). We also considered a dummy variable for being in the London NHS Region, but found no effect, so have excluded it in this paper.

Waiting Times: In addition, we add inpatient waiting times in the form of time between decision to admit and admission. (We use the raw values in the trees, but logged values in the regression modelling due to the highly non–normal distribution of the data.) McIntosh et al. (2012) find that higher waiting times are significantly associated with higher cancellation rates in NHS trusts.

Finally, we add year dummies to capture year–specific systematic effects common to all trusts which may affect cancellation rates.

Panel Regression Trees

While the main objective of this paper is to determine what 'predicts' a trust's cancellation rate, our aim is to portray possible alignments of management practices that may affect cancellations in a simple, easy-to-understand and intuitive way. We are interested in making meaningful sense of how several predictors (management practices) may be involved in complex interactions with one another when helping to explain trust efficiency. For example, we would like to understand whether the cancellation rates of trusts with a comparatively higher proportion of medical staff *and* greater workplace flexibility are systematically different from the cancellation rates observed in trusts with either a lower proportion of medical staff *or* lower workplace flexibility. Traditional parametric methods such as panel regression models do not always offer straightforward interpretation of such intricate interplay of variables. We therefore resort to a class of non-parametric techniques, commonly known in the machine learning literature (e.g., James et al. 2013) as regression trees that serve the purpose well. The tree mechanism involves recursively partitioning the predictor space into a number of smaller regions based on simple rules and then using the mean or median of the realised values (e.g., cancellation rate) of observations (e.g., trusts) belonging to an input region as the predicted value for a new observation that falls in that particular region. Most importantly, the splitting decision rules, order of importance of selected predictors and their interactions are summarised in a visually attractive and intuitive way. To our knowledge, this is one of the very few studies that exploits regression trees with an aim to identify optimal alignments of factors that contribute to trust efficiency and, in this case, minimises cancellation rates.

Several tree growing algorithms have been developed in the statistics and machine learning literature. They include CART (Breiman et al., 1984) and C4.5 (Quinlan, 2014) which function by maximising a statistical criterion over all possible predictors and split points simultaneously. These methods are often criticised for biased selection of variables which have many possible splits and missing values (Hothorn et al., 2006). In this paper, we opt to use a conditional inference framework proposed in Hothorn et al. (2006) which rectifies the problem of selection bias by choosing predictors for splitting based on a series of tests identifying statistically significant association between the predictors and the response variable.

Our data form a panel. Such a data structure requires careful accounting for possible variations across subjects that cannot be captured by observed predictors and also autocorrelation across observations from the same subject. If we observe subjects (e.g., trusts) $i = 1, 2, \ldots, I$ at times t (years) $= 1, 2, \ldots, T$ a general additive model can be defined as:

$$y_{it} = f(\mathbf{x}_{it}) + a_i + u_{it} \tag{1}$$

$$\begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix} \sim N(0, R_i) \tag{2}$$

$$a_i \sim N(0, D) \tag{3}$$

where for each subject i and time period t, y_{it} denotes the response of interest (here the cancellation rate), $\mathbf{x}_{it} = (x_{it1}, \ldots, x_{itK})'$ denotes a vector of K predictors (e.g., trust characteristics, regional features etc.) and u_{it} denote the errors. The a_i are the time-invariant, subject (here trust) –specific unobserved heterogeneity component. We allow for serial correlation in errors u_{it} from a particular subject by defining their covariance matrix R_i to be non-diagonal. The errors are, however, assumed to be independent across subjects and uncorrelated with a_i.

We seek to apply flexible tree-based methods for approximating the unknown relationship f which may well be non-linear. It is only over the last decade that progress has been made in generalising trees for panel data applications. This paper uses a method called Unbiased Random Effect Expectation Maximization (Unbiased RE-EM) Tree, recently developed by Fu and Simonoff (2015). Unlike many existing methods, including the standard RE-EM tree of Sela and Simonoff (2012), which rely on a CART type tree building algorithm, the Unbiased RE-EM Tree incorporates the conditional inference framework of Hothorn et al. (2006) and is, therefore, unbiased in nominating predictors for splitting. The Unbiased RE-EM Tree operates by alternating between two principal steps: one estimating f by using a tree method and the other estimating unobserved heterogeneities a_i by utilising a linear panel regression model. The process is initialised by setting the starting values of a_i to zero.

As with any non–parametric estimator, panel regression trees are subject to over–fitting. We estimate out–of–sample predictive accuracy of various 'plausible' models using cross–validation to achieve an optimal trade-off between the bias and variance (over–fitting). We select the tree model with the lowest prediction error to identify possible clusters of complementary management practices in the data. Our use of the panel regression tree technique is solely exploratory. The statistical significance of the clusters are then tested using panel regression techniques.

Exploring Management Practices

The regression tree technique arbitrarily selects from among highly correlated variables to build a model that best predicts the outcome variable (Kuhn and Johnson, 2013). In using the technique as an exploratory tool, one is required to classify the variables *in advance* into groups of uncorrelated variables and apply the method to each group to better understand the data.

To this end, we *a priori* select from among the correlated variables to build a base model, and then further examine the effect of replacing the explanatory variables with excluded correlated variables to assess the robustness of the results. A pair of variables is considered to be highly correlated when their correlation exceeds the threshold $\pm.80$ (Hinkle et al., 2003). We fit Unbiased RE-EM Tree models using a random intercept model to allow for variations among trusts due to unobserved attributes.

Among the management variables, the correlation between *Decisions* and *Communication* exceeds the threshold. We exclude *Communication* from our base model. This gives rise to the following model:

Tree Model I: *Cancellation Rate* $\sim$ *Team Quality* + *Intention to Leave* + *Flexible* + *Job Satisfaction* + *Appraisal* + *Ideas* + *Decisions* + *Job Design* + *Feedback* + *Errors* + *Supervisor*

There are further decisions involved in building a regression tree model such as deciding on the minimum number of observations in each terminal node, maximum number of layers in the tree and the significance level for splitting a node, which determine the complexity of the tree structure. A common approach is to rely on the out–of–sample predictive accuracy of alternative tree models to decide on the optimal tree structure. For exploratory purposes, we set a significance threshold of 10% for the unbiased RE-EM trees (i.e., a 10% false–positive rate for statistical significance is accepted). Setting the number of layers to 4 or 5 and the minimum number of observations to 10 or 20 generates sets of four possible models. Table 3 reports the root mean squared prediction error (RMSPE) of the models, calculated using 10–fold cross-validation. Figure 3 presents the unbiased regression tree with lowest RMSPE (0.387) or, in other words, the 'optimal' tree.

The tree presents a complex splitting of the sample into several segments, suggesting possible interactions among the variables. *Ideas* appears in the initial split of the tree, suggesting that effective management of ideas is the most important predictor of cancellation rates. Trusts whose score on *Ideas* exceeds 37 (i.e. a PRR of 37%) enjoy an overall lower cancellation rate: 0.8% versus 0.93%. (These figures are obtained by using the tree to segment the sample and calculate the mean of the response variable (Cancellation Rate) in each segment.) *Feedback* and *Appraisal* appear in the next layer as the second most predictively–significant variables. On the right–hand–side branch, where *Feedback* exceeds 37, the mean cancellation rate is 0.63% versus 0.83% for the other branch where *Feedback* falls below or is equal to 37. Several alignments of management practices appear as dominant. Strong management of *Ideas* ($>$ 37), *Feedback* ($>$ 37) and *Job Design* ($>$ 3.53) *jointly* yield quite a low mean cancellation rate 0.589%. The alignment where *Ideas* and *Appraisal* both fall below 37 and *Errors* lies below 3.4 gives rise to a relatively high mean cancellation rate (1.02%). The tree also reveals that higher levels of *Flexible* workplace or *Job Satisfaction* may give rise to higher mean cancellation rates. *Intention-to-Leave* fails to appear in the tree, so is not predictively significant.

It is likely that management practices interact with other trust features in driving performance. To capture this, we control for trust characteristics *Teaching Trust* status, *Foundation Trust* status and number of *Hospital Beds*. We also control for workforce skill–mix (proportion of medical staff (*Prop.Medical*), nurses (*Prop.Nurse*) and support staff (*Prop.Support*)), patient characteristics (proportions of emergencies (*Prop.Emergency*), female patients (*Prop.Female*) and *Patients aged above 60*)), and *Inpatient Waiting* time and *MFF*. Adding these variables to the management variables yields the model:

Tree Model II: *Cancellation Rate* $\sim$ *Teaching + FT + Hospital Beds + Prop. Medical + Prop. Nurse + Prop. Support + Prop. Emergency + Prop. Female + Patients aged above 60 + Inpatient Waiting + MFF + Team Quality + Intention to Leave + Flexible + Job Satisfaction + Appraisal + Ideas + Decisions + Job Design + Feedback + Errors + Supervisor*

Table 3 reports the RMSPE scores for the corresponding four models with the controls. We select the model with the lowest RMSPE (0.386). Figure 4 presents the unbiased tree. *Ideas* still occupies the initial node as predictively the most–significant variable, suggesting a key role for this management practice. The highest mean cancellation rate, 1.299%, occurs where trusts are not *Foundation Trusts* where *Ideas* is below or equal to 32, *Appraisal* is below or equal to 23, and *Inpatient Waiting* time is equal to or exceeds 3.912 days. In the branch where *Ideas* takes a value greater than 32, factors such as *Job Design* and *Prop.Medical* staff appear as predictively dominant whereas in trusts where *Ideas* takes a value of or below 32, practices such as *Appraisal* or *Supervisor* appear as predictively dominant. Both trees offer evidence on possible effects of management practices on cancellation rates. Even after controlling for a rich set of relevant covariates, management variables continue to retain their predictive significance and dominant positions in the tree.

Our goal is to use the tree models purely as exploratory tools to identify possible complementarities among management variables in driving cancellation rates. This demands ensuring that the tree results are stable. One can trust the structures in the trees only if they remain robust to (slight) variations in the data. This necessitates examining the impact on the tree structure of removing influential observations or carrying out the analysis on alternative randomly–selected subsets of the data. To test the stability of our results we have carried out a series of robustness checks. Removing a possibly influential observation (with cancellation rate 3.47% – See Figure 1) alters the trees only slightly. The key patterns remain valid. Adding an additional indicator variable for London to the controls to capture further regional differences did not change the tree structure. Our ultimate robustness check is to test the statistical significance of the observed complementarities (alignments of management practices) suggested by the trees using panel regression techniques. Diagnostic examination of the errors reasonably-well supports the normality assumption underlying the panel tree estimator (see the online supplement for further details).

Testing Complementarities

Complementarity of management practices suggests individual practices may not have a significant positive impact on efficiency, or may even have a negative impact. Testing the statistical significance of individual practices may yield misleading results. Technically, in testing management practices, one should consider testing the impacts of clusters of management practices rather than individual practices.

> Most of the problems of analyzing survey data have been reasonably well handled, except those revolving around the existence of interaction effects. ... Where interaction effects exist, the concept of a main effect is meaningless, and it is our belief that in human behavior there are so many interaction effects that we must change our approach to the problems of analysis (Morgan and Sonquist, 1963, pp.415 & 416).

Ichniowski et al. (1997) suggest testing for complementarity by introducing management practices multiplicatively into the panel regression model. That is,

$$y_{it} = c + \beta_1 m^1_{it} + \beta_2 m^2_{it} + \beta_{12}(m^1_{it} \times m^2_{it}) + \lambda \boldsymbol{x}_{it} + c_i + \tau_t + \epsilon_{it} \tag{4}$$

where m^1 and m^2 represent two management practices, whose relationship with productivity is such that output y_{it} increases by more when they are used together. $\boldsymbol{x}_{it}$ is a set of controls, c_i represents individual unobserved fixed effects and τ_t time effects. The complementarity hypothesis, in its simplest form, implies $\beta_{12} > 0$. A stronger implication is that the disruption caused by just using one practice alone actually could reduce productivity, i.e., $\beta_1 < 0$, $\beta_2 < 0$. In this case, a failure to include the interaction term in the regression specification may give rise to a zero coefficient for the individual practices (Bloom and Van Reenen, 2011).

A limitation of this approach is perhaps that when there are several management variables being considered, it becomes increasingly difficult to interpret the interaction coefficients. Further, with more management practices, the number of parameters being estimated increases substantially, reducing degrees of freedom (Mullahy, 2008). Equally important, management variables are often highly correlated, which can give rise to multicollinearity issues when they are jointly included in a regression model.

A second approach is to rely on the patterns in the trees to construct (dummy) indicator variables that show whether a trust does or does not exhibit a cluster of practices. The significance of the cluster can be tested by adding the dummy variable I to the control variables. That is to test a model of the form:

$$y_{it} = \alpha + \beta_i \mathbf{x}_{it} + \eta I + c_i + \tau_t + \epsilon_{it} \tag{5}$$

Since $I = 1$ when a trust meets the conditions defining the cluster, and $I = 0$ otherwise, the parameter η gives the difference in performance (cancellation rate) between the sub–set of trusts with the cluster of practices and those without, *given the same values for the control variables* $\mathbf{x}_{it}$*, the same* τ_t *(and the same error term* ϵ_{it}*).* Thus, the coefficient η determines whether the cluster makes any difference to trust performance (Wooldridge, 2013).

A third approach to testing complementarity builds on the idea that management practices impact trust efficiency by making production inputs, including human capital, more effective. As a consequence, the productivity of production inputs, specifically human capital, will differ under different alignments of management practices. This means it is possible to test the complementarity of management practices by interacting relevant indicator variables representing various alignments of practices with measures of workforce skill–mix or other inputs to estimate the impact of the clusters of practices. If the clusters have any impact on performance, the coefficients of the interaction terms will be statistically significant. The idea is to test factor models of the form:

$$y_{it} = \alpha + \beta_i \mathbf{x}_{it} + \gamma_i z_{it-1} + \eta_i \times I \times z_{it-1} + c_i + \tau_t + \epsilon_{it} \tag{6}$$

where $\mathbf{x}_{it}$ are control variables, z_{it} is the variable of interest (e.g., *Prop.Medical* staff), and I

refers to the indicator variable representing a cluster of management practices. η_i reflects the marginal effects of z_{it} in the sub–set of trusts where the indicator variable takes value one. The total effect of z_{it} for the cluster equals the sum $\gamma_i + \eta_i$ (Asker et al., 2014).

Although we report several results using the second approach, our analysis will mainly build on the third proposal. This approach allows testing how shifting from one cluster of practices to another would tend to alter the effect of input factors such as human capital (e.g. workforce skill-mix) on a trust's cancellation rate.

There are two common approaches to panel regression: Fixed Effects (FE) and Random Effects (RE) estimation. The RE estimator treats trusts' unobserved heterogeneity or characteristics c_i as a component of the independent error term ϵ_{it} (i.e., $Cov(x_{it}, c_i) = Cov(z_{it}, c_i) = 0$ for all $t = 1, 2, ..., T$ (Wooldridge, 2013)). There is generally no *a priori* grounds to warrant such an assumption. FE applies to situations where c_i is correlated with the explanatory variables x_{it} and z_{it}. The technique adopts a 'within estimation' specification using time demeaning to remove the unobserved heterogeneity c_i.

The FE estimator requires considerable 'within–unit' variation over time to estimate the coefficients consistently. Without adequate within–unit variation, the estimator is likely to generate large standard errors and suppresses the explanatory power of "slow–moving" variables of interest. The method also drops out time–invariant variables such as *Teaching Trust* and *Foundation Trust* status. Our explanatory variables are also persistent over time, and we are interested in estimating the effects of time–invariant trust characteristics. For these reasons, we turn to the Correlated Random Effects (CRE) technique, which can be traced back to the works of Chamberlain (1982) and Mundlak (1978) and has been further developed in Wooldridge (2013). This technique extends the RE approach and is comparable to the FE estimator. It also works well with unbalanced panel data.

As in Goerke and Pannenberg (2011), it is plausible to assume, as an approximation, that $Cov(x_{it}, c_i) = 0$ holds for the control variables. For example, c_i and patient demographics are independent. With this assumption, we focus on variables of interest, z_{it}, such as *Prop.Medical* staff. In CRE, we model correlation between c_i and $\{z_{it} : t = 1, 2, ..., T\}$. Since c_i is, by definition, constant over time, it can be correlated with the average level of the z_{it}. More specifically, let $\bar{z}_i = T^{-1}\Sigma_{i=1}^{t} z_{it}$ be the time average for trust i. It is plausible to assume the linear relationship

$$c_i = \alpha^* + \lambda \bar{z}_i + r_i \tag{7}$$

where r_i, by assumption, is uncorrelated with each z_{it}. The CRE approach uses (7) in conjunction with (6). Substituting the former into the latter gives:

$$y_{it} = \alpha + \beta_i \mathbf{x}_{it} + \gamma_i z_{it-1} + \eta_i \times I \times z_{it-1} + \lambda \bar{z}_i + r_i + \tau_t + \epsilon_{it} \tag{8}$$

The equation has a composite error term $r_i + \epsilon_{it}$, consisting of a time–invariant unobservable r_i and the idiosyncratic random noise ϵ_{it}. Further, because ϵ_{it} is assumed to be uncorrelated with z_{it}, for all i and t, ϵ_{it} is uncorrelated with $\bar{z}_i$. With these assumptions, the estimation problem becomes identical to the RE estimation of

$$y_{it} = \alpha + \beta_i \mathbf{x}_{it} + \gamma_i z_{it-1} + \eta_i \times I \times z_{it-1} + \lambda \bar{z}_i + \tau_t + u_{it}. \tag{9}$$

The addition of $\bar{z}_i$ controls for the correlation between c_i and the sequence $\{z_{it} : t = 1, 2, ..., T\}$. Wooldridge (2013) shows (chapter 10) that $\hat{\gamma}_{CRE} = \hat{\gamma}_{FE}$, where $\hat{\gamma}_{FE}$ is the FE estimator. $\hat{\eta}_i$ also represents an FE coefficient. Thus, the CRE approach provides a way to include time–invariant and slow–moving explanatory variables in what is effectively an FE analysis.

Results: Management Practices Matter

Our study pursues three aims: 1) discover potential alignments (clusters) of management practices present in the data; 2) test the statistical significance of each cluster; and 3) identify clusters that lower cancellation rates (increases efficiency) the most. The first unbiased regression tree (Figure 3) serves the first aim. The tree points to several alignments of management practices. Each alignment segments the sample into two subsets with different mean cancellation rates. As a first step, we test the significance of the difference by constructing a binary variable that takes value one if the alignment being considered holds for a trust and otherwise is zero. In testing the statistical significance of each cluster, we also include the controls in the model.

Table 4 reports the results of the exercise for several alignments. Column (1) adds an indicator variable (*idd1*) that takes value one whenever the within-trust means of *Ideas* and *Feedback* both exceed 37 and otherwise is zero. Column (2) adds an indicator variable (*idd2*) that takes value one whenever the within-trust means of *Ideas* and *Feedback* are greater than 37 and the within-trust mean of *Job design* exceeds 3.53. The indicator variable in Column (3) (*idd3*) takes value one whenever the within–trust mean of *Ideas* is greater than 37, *Feedback* falls below 37, *Flexible* lies below 67 and *Team Quality* exceeds 73. Finally, the indicator variable in Column (4) (*idd4*) takes value one whenever the within–trust mean of both *Ideas* and *Appraisal* exceed 37. The indicator variables are all significant at 5% or below, with a negative sign, i.e., all the clusters give rise to lower cancellation rates. The cluster captured by *idd1* in Column (1) has the largest coefficient (-0.271), so reduces cancellation rates the most. Shifting to the alignment where *Ideas* and *Feedback* are both scored above 37 tends to reduce cancellation rates by 0.271%. The regression tree, however, presents a richer set of alignments (clusters) of management practices and further analysis is needed to arrive at an optimal alignment. There could always be several well–aligned clusters of practices that equally enhance efficiency and lower cancellation rates. Moreover, in Table 4, trusts that have *Foundation Trust* status are consistently associated with lower rates of cancellation compared to those that don't, with effects from -0.069% to -0.086%. *Inpatient Waiting* times are consistently associated with higher cancellation rates and the relationship is significant at 1%. A one percentage–point increase in waiting times is associated with a 0.23% to 0.246% increase in cancellation rates.

The above approach does not indicate the channels through which management practices may impact efficiency. Management practices are likely to affect efficiency by making human capital or other inputs more productive. As a consequence, the productivity of a trust's human capital across various alignments of management practices will be different. Another approach to identifying an optimal cluster of practices, as discussed above, is to examine the productivity (efficiency) of human capital or other production inputs under different clusters. To this end, we interact relevant dummy variables with the measures of medical staff, nurses and support staff.

Consider the alignment of management practices where within-trust means of *Ideas* exceeds 37, *Feedback* exceeds 37 and *Job Design* exceeds 3.53, which was captured by *idd1*. We interact this alignment with our measures of human capital. Table 5 reports the results. Column (1) interacts the indicator variable with *Prop.Medical* staff, Column (2) interacts the indicator with *Prop.Nurse* staff, Column (3) interacts it with *Prop.Support* staff and Column (4) interacts it with the inpatient waiting measure. The interaction terms are all significant at 5% or below, with the expected negative signs. Thus, management practices affect efficiency by rendering human capital more productive. They appear particularly important for medical staff, as the resultant interaction has the largest of the interaction coefficients (-0.021).

Similarly, effective management may moderate impacts of factors, such as inpatient waits, that may increase the cancellation rate of elective operations. Column (4) interacts the dummy variable (*inter4*) with *Inpatient Waiting* time. The coefficient is -0.066 and significant at 1%. In trusts where the well–aligned cluster of management practices is in place, the adverse impact of inpatient waiting time on the cancellation rate is 0.36 (= 0.426 – 0.066) rather than 0.426 and the cancellation rate is lower by 0.021% due to medical staff being more effective. Similar to the previous result in Table 4, trusts with foundation trust status and lower inpatient waiting times are consistently and significantly associated with lower cancellation rates.

The first tree model (Model I, Figure 3) also suggests trusts where management of *Ideas* and *Appraisal* are both below the thresholds of 37 on average experience a higher cancellation rate. Using this information, it is possible to define a dummy variable to indicate membership of the subset of trusts with the alignment of practices where the within–trust means of both *Ideas* and *Appraisal* exceed 37. Interacting this indicator variable with measures of human capital and inpatient wait yields Table 6. The coefficients of the interaction terms (*inter1* to *inter4*) are all statistically significant at 1% or below with the expected negative signs.

The tree structure (Model I, Figure 3) reveals further information useful for identifying richer clusters. Consider *Flexible* workplace. It only appears in the path where *Ideas* exceeds 37 but *Feedback* falls below 37. Counter-intuitively, higher levels of workplace flexibility predict higher (worse) cancellation rates. This might be because *Flexible* workplace, *Ideas* and *Feedback* are complementary. That is, higher workplace flexibility will lower cancellation rates when complemented with effective management of ideas and feedback. The same consideration applies to *Job Satisfaction* and *Decisions*, where the tree predicts higher values will raise cancellations (i.e. lower efficiency). Again this might be due to the complementarity between these practices and *Ideas*. With effective management of *Ideas*, both *Decisions* and *Job Satisfaction* give rise to higher efficiency. Table 7 investigates these themes using several indicator variables. The indicator variable in Column (1) takes value one whenever the within–trust means of *Ideas* and *Feedback* exceed 37 but *Flexible* falls below its median (66.6) and otherwise is zero. The dummy variable in Column (2) is the same except the threshold on *Flexible* now is equal or above its median (66.6). We interact these binary variables with the proportion of medical staff. In column (1), the coefficient of the interaction term is -0.017 and in column (2) is -0.021, both significant at 1%. In the right cluster, higher levels of workplace flexibility give rise to lower cancellation rates. The indicator variable in column (3) takes value one whenever the within-trust means of *Ideas* exceeds 37, *Decision* lies above 28 and *Job Satisfaction* falls below its median 58.75. The dummy variable in column (4) is the same except that *Job Satisfaction* is now either equal to or exceeds its median 58.75. We interact both variables with the proportion of medical staff. In column (3), the interaction term is -0.011 significant at 10% and in column (4) is -0.016, significant at 1%. When complemented with effective management of decisions and ideas, higher levels of job satisfaction give rise to lower cancellation rates. Repeating the analysis for the *Prop.Nurse* staff, *Prop.Support* support staff and *Inpatient Waiting* time yields a similar of pattern of coefficients, with the magnitude of the coefficients being strongest for inpatient waiting time. Thus both workplace flexibility and job satisfaction lead to lower cancellation rates in the right alignments of management practices. Whether a management practice (e.g., workplace flexibility) lowers cancellation rates largely depends on other practices that are in place and their intensity.

While in carrying out the statistical tests we have precisely followed the splits in the first unbiased regression tree, the results are not entirely sensitive to the specific split–values. Varying the split–values, within 'reasonable' ranges, will leave the results unchanged. Finally, to illustrate this, we build on the 'general' information in the tree to construct an indicator variable that captures an alternative possible well–aligned cluster of complementary management practices to investigate the robustness of the results. The indicator variable takes value one whenever *Ideas*, *Decisions*, *Feedback*, *Job Design*, *Team Quality*, *Job Satisfaction*, *Appraisal* and *Supervisor* are jointly equal to or exceed their within–trust upper quartile values and otherwise is zero. We interact this indicator variable with the measures of human capital in our sample as well as with the measure of inpatient waiting time. Table 8 reports the results. The coefficients of the interaction terms are all negative and significant at the 1% level, with some having a greater absolute value than in previous tables. This cluster improves the efficiency of the medical staff and mitigates the adverse impact of inpatient waiting time slightly more strongly than the previous clusters.

Our analysis, so far, relied on the panel regression tree technique to identify potential interactions in the data. As an alternative approach, we explored the data using cluster analysis to identify clusters of complementary management practices in the data, and then tested the statistical significance of these clusters. The analysis yielded very similar results (see the online supplement). The online supplementary file also provides diagnostic plots on the empirical adequacy of the assumptions underlying the unbiased RE-EM tree models. The assumptions appear to be reasonable. Regardless of the choice of the technique, the panel regression models confirm the statistical significance of the interactions in the trees, provide evidence on the role of management practices in driving cancellation rates, and suggest ways to align management practices to reduce cancellation rates. The explanatory power of these models, measured using R^2 as shown in Tables 4–8, is not very high. There might be other factors that drive trust efficiency. Methodologically, the exercise reveals how regression trees can be used as a powerful exploratory tool for finding well–aligned clusters of complementary management practices. The economic magnitudes of the statistically significant coefficients of the interaction terms can guide identifying clusters or alignments of practices that most contribute to efficiency.

Discussion

Our analysis offers evidence on the impact of management practices on cancellation rates across trusts. Importantly, we find that different alignments of management practices give rise to quite different cancellation rates across trusts, with some being substantially lower. These results are well in line with the findings of Bloom and Van Reenen (2007), Bloom et al. (2014) and Bloom et al. (2015), who find that capital intensity or technology cannot alone fully account for large differences in Total Factor Productivity across firms or hospital organisations. Management practices and organisational design are most fundamental in driving hospital performance and quality. High management quality enhances performance, productivity and quality.

Our results are consistent with the hypothesis that management practices matter for hospital organisations' performance and, more importantly, it is the alignment of management practices that determine whether a practice negatively or positively affects performance. Management practices in isolation may not produce the desired effects or they may be counterproductive (Ichniowski et al., 1997; Vermeeren et al., 2014; West et al., 2006). *Workplace Flexibility* gives rise to lower cancellation rates only when management practices such as *Ideas* and *Feedback* receive high scores. When *Feedback* receives a relatively low score, *Workplace Flexibility* is associated with higher short-notice cancellation rates. Management practices also strongly complement the skill-mix variables such as the *Proportion of Medical Staff*. The online supplementary file documents further complementarities. Management practices such as *Decisions*, *Team Quality* and *Feedback* have higher impact on performance when other management practices are in place at a high level. Indeed, individual practices often fail to appear as statistically significant. Thus implementing organisational change calls for a holistic approach, requiring a shift from one alignment of management practices to another. This could help explain why imitating good practices often proves challenging. The focus of policy should be on clusters of complementary practices. Any policy effort to lower cancellation rates across hospitals should involve understanding sources of heterogeneity in clusters of management practices and deal with the causes. In addition, we find that trusts without foundation trust status and with longer inpatient waiting times are consistently and significantly associated with higher cancellation rates.

As a methodological contribution, our approach suggests a powerful method for systematically searching for optimal alignments of management practices, contributing to the literature on the complementarity of organisational practices. This is independently valuable as the management science literature currently lacks techniques for effectively searching for optimal alignments of management practices. In the absence of theory, one is largely left with trial and error.

There are several caveats to our study. Our panel is fairly short and both the dependent and management variables are slow moving. Consequently, the data may lack adequate variation, which can limit the reliability of our results. With a longer dataset, there will be more variation in the data and one can run fixed effects regressions effectively. A second limitation relates to the measurement of the management practices. While the NHS staff survey has a response rate of around 50% on average over the years (Pinder et al., 2013), there can be legitimate questions on how adequately the responses capture the quality of practices in the trusts. Finally, while efforts are made centrally in the NHS to ensure the quality of the datasets, it may contain noise, as does data from any real complex system. This could limit the strength of our analysis. Proudlove et al. (2018a) suggest, though, that there are robust signals of system performance in similar public NHS data.

The observed variation in management practices raises the deeper question of why management practices differ materially between hospital organisations. This question will shape our future research. We also aim to further our understanding of complementarities between management practices and other organisational routines, which may affect the efficiency of patient treatment.

References

Ahmad, S. and Schroeder, R. G. (2003). The impact of human resource management practices on operational performance: recognizing country and industry differences. *Journal of*

Operations Management, 21(1):19–43.

Ali, M., Salehnejad, R., and Mansur, M. (2018a). Hospital heterogeneity: what drives the quality of health care. *The European Journal of Health Economics*, 19(3):385–408.

Ali, M., Salehnejad, R., and Mansur, M. (2018b). Hospital productivity: the role of efficiency drivers. *International Journal of Health Planning and Management*, in press.

Allen, P., Wright, J., Keen, J., Dempster, P., Hutchings, A., Townsend, J., Street, A., and Verzulli, R. (2010). Investigating the Governance of NHS Foundation Trusts, Final Project Report. Technical report, NIHR SDO Programme.

Appelbaum, E., Bailey, T., Berg, P., and Kalleberg, A. L. (2000). *Manufacturing Advantage: Why High-Performance Work Systems Pay Off*. Cornell University Press, Ithaca, NY.

Argo, J. L., Vick, C. C., Graham, L. A., Itani, K. M., Bishop, M. J., and Hawn, M. T. (2009). Elective Surgical Case Cancellation in the Veterans Health Administration System: Identifying Areas for Improvement. *The American Journal of Surgery*, 198(5):600–606.

Asker, J., Farre-Mensa, J., and Ljungqvist, A. (2014). Corporate Investment and Stock Market Listing: A puzzle? *The Review of Financial Studies*, 28(2):342–390.

Basson, M. D., Butler, T. W., and Verma, H. (2006). Predicting Patient Non–appearance for Surgery as a Scheduling Strategy to Optimize Operating Room Utilization in a Veterans' Administration hospital. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 104(4):826–834.

Birdi, K., Clegg, C., Patterson, M., Robinson, A., Stride, C. B., Wall, T. D., and Wood, S. J. (2008). The impact of human resource and operational management practices on company productivity: A longitudinal study. *Personnel Psychology*, 61(3):467–501.

Bloom, N., Lemos, R., Sadun, R., Scur, D., and Reenen, J. V. (2014). The New Empirical Economics of Management. *CEP Occasional Papers* 41, Centre for Economic Performance, London School of Economics.

Bloom, N., Propper, C., Seiler, S., and Van Reenen, J. (2015). The impact of Competition on Management Quality: Evidence from Public Hospitals. *The Review of Economic Studies*, 82(2):457–489.

Bloom, N., Sadun, R., and Van Reenen, J. (2013). Does Management Matter in Healthcare. Technical report, London School of Economics Mimeo.

Bloom, N. and Van Reenen, J. (2007). Measuring and Explaining Management Practices Across Firms and Countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.

Bloom, N. and Van Reenen, J. (2011). Human Resource Management and Productivity. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, chapter 19, pages 1697–1767. Elsevier, Amsterdam.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. CRC Press, New York.

Brynjolfsson, E. and Milgrom, P. (2013). Complementarity in organizations. In Gibbons, R. and Roberts, J., editors, *The handbook of organizational economics*, pages 11–55. Princeton University Press, Princeton.

Bryson, A., Forth, J., and Stokes, L. (2017). Does employees' subjective well-being affect workplace performance? *Human Relations*, 70(8):1017–1037.

Carter, P. (2016). *Operational Productivity and Performance in English NHS Acute Hospitals: Unwarranted Variations. An independent Report for the Department of Health by Lord Carter of Coles*. Department of Health, London. `https://www.gov.uk/government/publications/productivity-in-nhs-hospitals`, accessed 5 Sept 2017.

Castelli, A., Street, A., Verzulli, R., and Ward, P. (2015). Examining Variations in Hospital Productivity in the English NHS. *The European Journal of Health Economics*, 16(3):243–254.

Chamberlain, G. (1982). Multivariate Regression Models for Panel Data. *Journal of Econometrics*, 18(1):5–46.

Chandra, A., Finkelstein, A., Sacarny, A., and Syverson, C. (2016). Productivity Dispersion in Medicine and Manufacturing. *The American Economic Review*, 106(5):99–103.

Chuang, E., Dill, J., Morgan, J. C., and Konrad, T. R. (2012). A configurational approach to the relationship between high–performance work practices and frontline health care worker outcomes. *Health Services Research*, 47(4):1460–1481.

Cookson, G., Jones, S., and Laliotis, I. (2017). Cancelled Procedures in the English NHS: Evidence from the 2010 Tariff Reform. *Health Economics*, 26:e126—e139.

Cookson, G., Jones, S., and McIntosh, B. (2013). Cancelled Procedures: Inequality, Inequity and the National Health Service Reforms. *Health Economics*, 22(7):870–876.

Department of Health (2011). Guidance manual for the NHS Staff Survey 2011. Technical report, NHS Staff Survey Co-ordination Centre. `http://doc.ukdataservice.ac.uk/doc/7203/mrdoc/pdf/7203_guidance_notes.pdf`, accessed 13 Nov 2018.

Dimitriadis, P., Iyer, S., and Evgeniou, E. (2013). The Challenge of Cancellations on the Day of Surgery. *International Journal of Surgery*, 11(10):1126–1130.

Fu, W. and Simonoff, J. S. (2015). Unbiased Regression Trees for Longitudinal and Clustered Data. *Computational Statistics & Data Analysis*, 88:53–74.

Glowacka, K. J., Henry, R. M., and May, J. H. (2009). A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling. *Journal of the Operational Research Society*, 60(8):1056–1068.

Goerke, L. and Pannenberg, M. (2011). Trade Union Membership and Dismissals. *Labour Economics*, 18(6):810–821.

Haana, V., Sethuraman, K., Stephens, L., Rosen, H., and Meara, J. G. (2009). Case Cancellations On The Day of Surgery: An Investigation in an Australian Paediatric Hospital. *ANZ Journal of Surgery*, 79(9):636–640.

Hand, R., Levin, P., and Stanziola, A. (1990). The Causes of Cancelled Elective Surgery. *Quality Assurance and Utilization Review*, 5(1):2–6.

Harper, P. R. and Shahani, A. (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, 53(1):11–18.

Harvey, G., Jas, P., Walshe, K., and Skelcher, C. (2015). Analysing Organisational Context: Case Studies on the Contribution of Absorptive Capacity Theory to Understanding Inter–Organisational Variation in Performance Improvement. *BMJ Quality & Safety*, 24(1):48–55.

Hinkle, D. E., Wiersma, W., and Jurs, S. G. (2003). *Applied statistics for the behavioral sciences*. Houghton Mifflin, Boston, 5th edition.

Holmstrom, B. and Milgrom, P. (1994). The Firm as an Incentive System. *The American Economic Review*, 84(4):972–991.

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674.

Hovlid, E., von Plessen, C., Haug, K., Aslaksen, A. B., and Bukve, O. (2013). Patient Experiences with Interventions to Reduce Surgery Cancellations: A Qualitative Study. *BMC Surgery*, 13(1):30.

Huang, Y.-L. and Marcak, J. (2013). Radiology scheduling with consideration of patient characteristics to improve patient access to care and medical resource utilization. *Health Systems*, 2(2):93–102.

Ichniowski, C., Shaw, K., and Prennushi, G. (1997). The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines. *The American Economic Review*, 87(3):291–313.

Ivarsson, B., Larsson, S., and Sjöberg, T. (2004). Postponed or Cancelled Heart Operations from the Patient's Perspective. *Journal of Nursing Management*, 12(1):28–36.

Ivarsson, B., Ola Kimblad, P., Sjöberg, T., and Larsson, S. (2002). Patient Reactions to Cancelled or Postponed Heart Operations. *Journal of Nursing Management*, 10(2):75–81.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer, New York.

Kaplan, H. C., Brady, P. W., Dritz, M. C., Hooper, D. K., Linam, W., Froehle, C. M., and Margolis, P. (2010). The Influence of Context on Quality Improvement Success in Health Care: A Systematic Review of the Literature. *The Milbank Quarterly*, 88(4):500–559.

Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*. Springer, New York.

Lacqua, M. J. and Evans, J. T. (1994). Cancelled Elective Surgery: An Evaluation. *The American Surgeon*, 60(11):809–811.

Lega, F., Prenestini, A., and Spurgeon, P. (2013). Is Management Essential to Improving the Performance and Sustainability of Health Care Systems and Organizations? A Systematic Review and a Roadmap for Future Studies. *Value in Health*, 16(1):S46–S51.

Mangan, J. L., Walsh, C., Kernohan, W., Murphy, J., Mollan, R., McMillen, R., and Beverland, D. (1992). Total Joint Replacement: Implication of Cancelled Operations for Hospital Costs and Waiting List Management. *Quality and Safety in Health Care*, 1(1):34–37.

McIntosh, B., Cookson, G., and Jones, S. (2012). Cancelled Surgeries and Payment by Results in the English National Health Service. *Journal of Health Services Research & Policy*, 17(2):79–86.

Milgrom, P. and Roberts, J. (1995). Complementarities and Fit Strategy, Structure, and Organizational Change in Manufacturing. *Journal of Accounting and Economics*, 19(2):179–208.

Morgan, J. N. and Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 58(302):415–434.

Mullahy, J. (2008). Interaction effects. Technical report, University of Wisconsin-Madison and NBER. `https://www.le.ac.uk/economics/healtheconometrics/documents/IAE2.pdf`, accessed 6 Sept 2017.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69–85.

Nasr, A., Reichardt, K., Fitzgerald, K., Arumugusamy, M., Keeling, P., and Walsh, T. (2004). Impact of Emergency Admissions on Elective Surgical Workload. *Irish Journal of Medical Science*, 173(3):133–135.

NHS England (2016). NHS cancelled elective operations: quarter ending March 2016. Technical report, NHS England. `https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/04/NHS-cancelled-elective-operations-commentary-Q4-2015-16.pdf`, accessed 6 Sept 2017.

O'Malley, N., Sproule, J., Nicholson, P., Rice, J., and McElwain, J. (2005). Ring–Fencing — Bed Protection for the Patient. *Irish Journal of Medical Science*, 174(Suppl 1):49–49.

Pinder, R. J., Greaves, F. E., Aylin, P. P., Jarman, B., and Bottle, A. (2013). Staff perceptions of quality of care: An observational study of the NHS Staff Survey in hospitals in England. *BMJ Quality & Safety*, 22(7):563–570.

Pollard, J. B. and Olson, L. (1999). Early outpatient preoperative anesthesia assessment: Does it help to reduce operating room cancellations? *Anesthesia & Analgesia*, 89(2):502–505.

Proudlove, N., Gordon, K., and Boaden, R. (2003). Can good bed management solve the overcrowding in Accident and Emergency departments? *Emergency Medicine Journal*, 20(2):149–155.

Proudlove, N. C., Goff, M., Walshe, K., and Boaden, R. (2018a). The signal in the noise: Robust detection of performance "outliers" in health services. *Journal of the Operational Research Society*, online first:1–13.

Proudlove, N. C., Samarasinghe, B. S., and Walshe, K. (2018b). Investigating consistent patterns of variation in short-notice cancellations of elective operations: The potential for learning and improvement through multi-site evaluations. *Health Services Management Research*, 31(3):111–119.

Quinlan, J. R. (2014). *C4.5: Programs for Machine Learning*. Elsevier.

Rai, M. and Pandit, J. (2003). Day of surgery cancellations after nurse–led pre–assessment in an elective surgical centre: The first 2 years. *Anaesthesia*, 58(7):692–699.

Robinson, S., Alifantis, T., Edwards, J. S., Ladbrook, J., and Waller, A. (2005). Knowledge-based improvement: Simulation and artificial intelligence for identifying and improving human decision-making in an operations system. *Journal of the Operational Research Society*, 56(8):912–921.

Rymaruk, S. (2011). A retrospective observational study of patient cancellations on the day of surgery in the general surgical directorate. *The Journal of Perioperative Practice*, 21(10):337–341.

Sanjay, P., Dodds, A., Miller, E., Arumugam, P., and Woodward, A. (2007). Cancelled elective operations: An observational study from a district general hospital. *Journal of Health Organization and Management*, 21(1):54–58.

Schofield, W. N., Rubin, G. L., Piza, M., Lai, Y. Y., Sindhusake, D., Fearnside, M. R., and Klineberg, P. L. (2005). Cancellation of Operations on the Day of Intended Surgery at a Major Australian Referral Hospital. *Medical Journal of Australia*, 182(12):612–615.

Schuster, M., Neumann, C., Neumann, K., Braun, J., Geldner, G., Martin, J., Spies, C., Bauer, M., Group, C. S., et al. (2011). The Effect of Hospital Size and Surgical Service on Case Cancellation in Elective Surgery: Results from a Prospective Multicenter Study. *Anesthesia & Analgesia*, 113(3):578–585.

Seim, A. R., Fagerhaug, T., Ryen, S. M., Curran, P., Sæther, O. D., Myhre, H. O., and Sandberg, W. S. (2009). Causes of Cancellations on the Day of Surgery at Two Major University Hospitals. *Surgical Innovation*, 16(2):173–180.

Sela, R. J. and Simonoff, J. S. (2012). RE–EM trees: A Data Mining Approach for Longitudinal and Clustered Data. *Machine Learning*, 86(2):169–207.

Singh, G. C., Agha, R., and Roberts, D. R. (2005). Cancellations in Day–Case ENT Surgery. *Ambulatory Surgery*, 12(2):57–60.

Tait, A. R., Voepel-Lewis, T., Munro, H. M., Gutstein, H. B., and Reynolds, P. I. (1997). Cancellation of Pediatric Outpatient Surgery: Economic and Emotional Implications for Patients and their Families. *Journal of Clinical Anesthesia*, 9(3):213–219.

Vermeeren, B., Kuipers, B., and Steijn, B. (2014). Does leadership style make a difference? linking hrm, job satisfaction, and organizational performance. *Review of Public Personnel Administration*, 34(2):174–195.

West, M. A., Guthrie, J. P., Dawson, J. F., Borrill, C. S., and Carter, M. (2006). Reducing patient mortality in hospitals: the role of human resource management. *Journal of Organizational Behavior*, 27(7):983–1002.

Wong, D., Harris, S., Moonesinghe, S., Moonesinghe, S. R., Wong, D. J., Harris, S. K., Bedford, J., Boney, O., Chazapis, M., Drake, S., et al. (2018). Cancelled operations: a 7-day cohort study of planned adult inpatient surgery in 245 UK National Health Service hospitals. *British Journal of Anaesthesia*, 121(4):730–738.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South Western, Mason, OH, 5th edition.

Table 1: Data Definitions and Sources

Data	*Definition*	*Source*
Dependent Variable		
Cancellation Rate (%)	Percentage of elective operations cancelled due to non–clinical reasons over total elective admissions	NHS England
Management, Culture and Environment	(All measured on a 0–100 scale unless indicated otherwise)	
Management: Decisions	PRR: % who answered agree or strongly agree to 'Senior managers here try to involve staff in important decisions'	NHS Staff Survey
Management: Feedback	PRR: % who answered agree or strongly agree to 'Senior managers act on staff feedback'	NHS Staff Survey
Management: Ideas	PRR: % who answered agree or strongly agree to 'Senior Managers encourage staff to suggest new ideas for improving services'	NHS Staff Survey*
Management: Communication	PRR: % who answered agree or strongly agree to 'Communication between senior management and staff is effective'	NHS Staff Survey
Flexible Working Practice	PRR: % using flexible working options (Key summary score from NHS Staff Summary)	NHS Staff Survey*
Job Satisfaction	Average positive response rates of 7 measures of staff satisfaction (Key summary score from NSS)	NHS Staff Survey
Job Design	Quality of job design. The scale assesses the extent to which staff are performing jobs that are well designed and rich in content (Score 1–5)	NHS Staff Survey*
Team Quality	Effective Team Working	NHS Staff Survey

Continued on next page...

Table 1: Data Definitions and Sources

Data	*Definition*	*Source*
Errors	Fairness and effectiveness of procedures for reporting errors, near misses and incidents given on a score of 1 to 5 where 5 represents a higher score	NHS Staff Survey
Appraisal	% having well structured appraisal reviews within previous 12 months	NHS Staff Survey
Supervisor	Support from immediate managers (Score 1–5)	NHS Staff Survey
Intention to Leave	The extent (%) to which staff are considering leaving their organisation, and looking for a new job either within or outside of the NHS	NHS Staff Survey*
Trust Characteristics		
Teaching Trust status	Equal to 1 if a teaching trust and 0 otherwise	NHS England
Foundation Trust (FT) status	Equal to 1 if has FT status and 0 otherwise	Monitor
Hospital Beds	Average number of available beds	NHS England
Workforce Skill–Mix		
Medical Staff (%)	(Medical workforce / Total Staff) $\times$ 100	NHS Workforce Statistics
Nursing Staff (%)	(Nursing workforce / Total Staff) $\times$ 100	NHS Workforce Statistics
Support Staff (%)	(Support Staff / Total Staff) $\times$ 100	NHS Workforce Statistics
Patient Characteristics		
Patients aged over 60 (%)	Proportion of patients (FCE) aged 60 years and older	NHS Digital
Proportion of Emergency patients (%)	(Emergency admissions / Total admissions) $\times$ 100	NHS Digital

Continued on next page...

Table 1: Data Definitions and Sources

Data	*Definition*	*Source*
Proportion of female patients (%)	(Female admissions / Total admissions) $\times$ 100	NHS Digital
Geographical Characteristics		
Market Forces Factor (MFF)	MFF $\times$ 100	Department of Health
Waiting Times		
Inpatient Waiting Times	The average waiting time (days) for admission from the waiting list for elective admissions of types Waiting List Admission and Booked Admission.	NHS Digital

*The reporting of this variable was discontinued after 2011. The data for 2012/13 were estimated using the mean of previous years

Table 2: Summary Statistics

Variables	*N*	*Mean*	*St. Dev.*	*Min*	*Max*
Dependent Variable					
Cancellation Rate (%)	578	0.857	0.372	0.000	3.470
Management, Culture and Environment					
Management: Decisions	577	25.731	5.439	8	46
Management: Feedback	577	28.438	5.459	7	47
Management: Ideas	576	37.264	6.697	16	61
Management: Communication	577	27.392	7.036	8	53
Flexible	576	65.061	5.312	47	82
Job Satisfaction	577	59.450	3.563	48.400	69.600
Job Design	576	3.395	0.068	3.180	3.600
Team Quality	576	74.746	2.414	65.500	81.000
Errors	575	3.457	0.086	3.170	3.710
Appraisal	577	33.324	6.095	17	52
Supervisor	577	3.604	0.090	3.260	3.880
Intention to Leave	575	21.61	2.856	14.750	31.00
Trust Characteristics					
Hospital Beds	575	758	346	222	2196
Workforce Skill–Mix					
Prop. Medical	576	11.379	2.249	4.674	19.820
Prop. Nurse	578	31.848	3.607	22.811	47.795

Continued on next page. . .

Table 2: Summary Statistics

Variables	*N*	*Mean*	*St. Dev.*	*Min*	*Max*
Prop. Support	578	29.175	4.058	11.932	39.057
Patient Characteristics					
Patients aged above 60	578	45.959	7.172	21.830	70.490
Prop. Emergency	578	37.361	5.658	15.948	61.343
Prop. Female	578	56.774	2.788	47.289	70.103
MFF & Inpatient Waiting Time					
MFF	577	108.454	7.065	100.222	132.074
Inpatient Waiting	576	49.234	15.172	0	344

Table 3: Tree Model Assessment

	alpha	layers	min Obs	CV Score 10
Tree Model I				
Model 1	0.1	4	10	0.387
Model 2	0.1	4	20	0.395
Model 3	0.1	5	10	0.394
Model 4	0.1	5	20	0.400
Tree Model II				
Model 1	0.1	4	10	0.392
Model 2	0.1	4	20	0.386
Model 3	0.1	5	10	0.391
Model 4	0.1	5	20	0.390

Table 4: Indicator Variable Models: Complementarity Results

	(1)	(2)	(3)	(4)
Intercept	0.300 (1.101)	0.372 (1.094)	0.184 (1.088)	0.325 (1.104)
Trust Characteristics				
Teaching	0.004 (0.076)	−0.015 (0.077)	−0.020 (0.079)	0.020 (0.075)
Foundation Trust	−0.076 (0.040)*	−0.086 (0.040)**	−0.082 (0.040)**	−0.069 (0.039)*
Hospital Beds	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Workforce Skill–Mix				
Prop. Medical	−0.008 (0.011)	−0.008 (0.011)	−0.007 (0.011)	−0.008 (0.011)
Prop. Nurse	0.007 (0.006)	0.007 (0.006)	0.007 (0.006)	0.007 (0.006)
Prop. Support	0.009 (0.007)	0.008 (0.007)	0.007 (0.007)	0.007 (0.007)
Patient Characteristics				
Prop. Emergency	−0.003 (0.004)	−0.003 (0.004)	−0.003 (0.004)	−0.003 (0.004)
Prop. Female	−0.013 (0.011)	−0.013 (0.011)	−0.012 (0.010)	−0.012 (0.010)
Patients aged above 60	−0.002 (0.006)	−0.001 (0.006)	−0.000 (0.006)	−0.001 (0.006)
MFF and Waiting Times				
Inpatient Waiting	0.244 (0.083)***	0.246 (0.082)***	0.245 (0.080)***	0.230 (0.083)***
MFF	0.002 (0.005)	0.001 (0.005)	0.002 (0.005)	0.002 (0.005)
Interactions				
idd1	−0.271 (0.073)***			
idd2		−0.252 (0.102)**		
idd3			−0.117 (0.056)**	
idd4				−0.263 (0.060)***
Year Dummies				
2010/11	−0.033 (0.041)	−0.035 (0.041)	−0.030 (0.041)	−0.031 (0.040)
2011/12	−0.087 (0.044)**	−0.089 (0.044)**	−0.087 (0.044)**	−0.088 (0.044)**
2012/13	−0.152 (0.045)***	−0.154 (0.045)***	−0.152 (0.045)***	−0.152 (0.045)***
R^2	0.100	0.100	0.100	0.105
Adj. R^2	0.100	0.092	0.100	0.102
Num. obs.	551	551	551	551

This table reports initial results on possible impact of clusters in the first regression tree by separately adding a number of indicator variables to the control variables. The binary variable *idd1* takes value one whenever the within–trust mean of *Ideas* and *Feedback* exceed 37 and otherwise zero. The binary variable *idd2* takes value one whenever the within–trust mean of *Ideas* and *Feedback* are greater than 37 and within–trust mean of *Job Design* lies above 3.53. The binary variable *idd3* takes value one for trust where the within–trust means of *Ideas* and *Feedback* exceed 37, *Flexible* lies below 67.60 and *Team Quality* exceeds 73.7. Finally, the binary variable *idd4* takes value one whenever the within–trust means of *Ideas* and *Appraisal* are both above 37 and otherwise zero. The skill–mix variables and *Waiting Time* are lagged by one year. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 5: Interaction Models: Ideas, Feedback & Job Design

	(1)	(2)	(3)	(4)
Intercept	0.156 (1.147)	0.970 (1.137)	0.167 (1.103)	0.087 (1.113)
Trust Characteristics				
Teaching	0.010 (0.084)	−0.025 (0.078)	−0.001 (0.080)	0.002 (0.077)
Foundation Trust	−0.089 (0.041)**	−0.113 (0.040)***	−0.089 (0.040)**	−0.103 (0.040)**
Hospital Beds	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Workforce Skill–Mix				
Prop. Medical	0.001 (0.013)	0.009 (0.012)	−0.008 (0.011)	−0.006 (0.011)
Prop. Nurse	0.009 (0.007)	0.028 (0.009)***	0.006 (0.006)	0.007 (0.006)
Prop. Support	0.008 (0.007)	0.013 (0.007)**	0.003 (0.010)	0.008 (0.007)
Patient Characteristics				
Prop. Emergency	−0.003 (0.004)	−0.003 (0.004)	−0.003 (0.004)	−0.002 (0.004)
Prop. Female	−0.012 (0.011)	−0.015 (0.011)	−0.014 (0.011)	−0.013 (0.010)
Patients aged above 60	−0.001 (0.006)	−0.006 (0.006)	−0.002 (0.006)	0.000 (0.006)
MFF and Waiting Times				
Inpatient Waiting	0.272 (0.085)***	0.265 (0.084)***	0.256 (0.081)***	0.426 (0.153)***
MFF	0.003 (0.006)	0.004 (0.005)	0.002 (0.005)	0.002 (0.005)
Averages				
Prop. Medical	−0.026 (0.027)			
Prop. Nurse		−0.052 (0.012)***		
Prop. Support			0.010 (0.015)	
Inpatient Waiting				−0.012 (0.005)***
Interactions				
inter1	−0.021 (0.008)***			
inter2		−0.006 (0.003)**		
inter3			−0.007 (0.004)**	
inter4				−0.066 (0.027)**
Year Dummies				
2010/11	−0.026 (0.043)	−0.019 (0.040)	−0.029 (0.043)	−0.022 (0.038)
2011/12	−0.085 (0.046)*	−0.048 (0.042)	−0.088 (0.045)**	−0.094 (0.045)**
2012/13	−0.147 (0.047)***	−0.113 (0.043)***	−0.152 (0.046)***	−0.161 (0.047)***
R^2	0.097	0.121	0.096	0.097
Adj. R^2	0.094	0.118	0.093	0.094
Num. obs.	545	551	551	548

This table reports the first set of results using the third approach to testing complementarity. We build an indicator variable that takes value one whenever *Ideas* exceeds 37, *Feedback* exceeds 37 and *Job Design* lies above 3.53. Column 1 interacts the indicator variable with the proportion of medical staff. Column 2 interacts the binary variable with the proportion of nurses. Column 3 interacts the variable with the proportion of support staff. Column 4 interacts the binary variable with the inpatient waiting time measure. All the results are CRE results. The variable of interest in each column appears twice – once in its original form and once as its within–trust mean. With the within–trust mean being included, the coefficient of the interaction term represents the fixed–effects estimate. The skill–mix variables and *Waiting Time* are lagged by one year. All the models include year dummies. Robust standard errors are given in parentheses.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 6: Interaction Models: Ideas and Appraisal

	(1)	(2)	(3)	(4)
Intercept	0.156 (1.156)	0.904 (1.144)	0.129 (1.112)	0.026 (1.121)
Trust Characteristics				
Teaching	0.041 (0.082)	0.007 (0.076)	0.033 (0.077)	0.041 (0.075)
Foundation Trust	-0.069 (0.040)*	-0.095 (0.040)**	-0.073 (0.039)*	-0.086 (0.039)**
Hospital Beds	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Workforce Skill–Mix				
Prop. Medical	0.001 (0.013)	0.009 (0.012)	-0.008 (0.011)	-0.005 (0.011)
Prop. Nurse	0.008 (0.007)	0.027 (0.009)***	0.006 (0.006)	0.006 (0.006)
Prop. Support	0.006 (0.007)	0.012 (0.006)*	0.004 (0.010)	0.006 (0.006)
Patient Characteristics				
Prop. Emergency	-0.004 (0.004)	-0.004 (0.004)	-0.003 (0.004)	-0.002 (0.004)
Prop. Female	-0.011 (0.011)	-0.014 (0.011)	-0.012 (0.011)	-0.011 (0.010)
Patients aged above 60	-0.002 (0.006)	-0.006 (0.006)	-0.001 (0.006)	-0.000 (0.006)
MFF and Waiting Times				
Inpatient Waiting	0.255 (0.086)***	0.249 (0.084)***	0.238 (0.081)***	0.432 (0.153)***
MFF	0.004 (0.006)	0.005 (0.005)	0.003 (0.005)	0.003 (0.005)
Averages				
Prop. Medical	-0.019 (0.026)			
Prop. Nurse		-0.051 (0.012)***		
Prop. Support			0.007 (0.015)	
Inpatient Waiting				-0.013 (0.005)***
Interactions				
inter1	-0.020 (0.005)***			
inter2		-0.007 (0.002)***		
inter3			-0.010 (0.002)***	
inter4				-0.077 (0.016)***
Year Dummies				
2010/11	-0.025 (0.042)	-0.016 (0.039)	-0.028 (0.042)	-0.018 (0.037)
2011/12	-0.086 (0.045)*	-0.048 (0.042)	-0.088 (0.044)**	-0.094 (0.045)**
2012/13	-0.149 (0.047)***	-0.113 (0.043)***	-0.152 (0.046)***	-0.161 (0.046)***
R^2	0.106	0.130	0.106	0.110
Adj. R^2	0.103	0.126	0.103	0.107
Num. obs.	545	551	551	548

We build an indicator variable that takes value one whenever both *Ideas* and *Appraisal* exceed 37 and otherwise zero. Column 1 interacts the indicator variable with the proportion of medical staff. Column 2 interacts the binary variable with the proportion of nurses. Column 3 interacts the variable with the proportion of support staff. Column 4 interacts the binary variable with the inpatient waiting time measure. All the results are CRE results. The variable of interest in each column appears twice – once in its original form and once as its within–trust mean. with the within–trust mean being included, the coefficient of the interaction term represents the fixed–effects estimate. The skill–mix variables and *Waiting Time* are lagged by one year. All the models include year dummies. And robust standard errors are given in parentheses.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 7: Interaction Models: Ideas, Feedback & Workplace Flexibility

	(1)	(2)	(3)	(4)
Intercept	0.107 (1.143)	0.129 (1.153)	0.064 (1.136)	0.292 (1.158)
Trust Characteristics				
Teaching	0.020 (0.084)	0.022 (0.083)	0.013 (0.084)	0.019 (0.084)
Foundation Trust	-0.086 (0.041)**	-0.086 (0.041)**	-0.097 (0.041)**	-0.058 (0.043)
Hospital Beds	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Workforce Skill–Mix				
Prop.Medical	0.001 (0.013)	0.002 (0.013)	0.001 (0.013)	0.003 (0.013)
Prop.Nurse	0.009 (0.007)	0.009 (0.007)	0.008 (0.007)	0.008 (0.007)
Prop.Support	0.008 (0.007)	0.009 (0.007)	0.008 (0.007)	0.005 (0.007)
Patient Characteristics				
Prop.Emergency	-0.003 (0.004)	-0.003 (0.004)	-0.003 (0.004)	-0.004 (0.004)
Prop.Female	-0.012 (0.011)	-0.013 (0.011)	-0.011 (0.011)	-0.013 (0.011)
Patients aged above 60	-0.001 (0.006)	-0.002 (0.006)	-0.001 (0.006)	-0.002 (0.006)
MFF and Waiting Time				
Inpatient Waiting	0.274 (0.085)***	0.269 (0.085)***	0.270 (0.085)***	0.270 (0.083)***
MFF	0.003 (0.006)	0.004 (0.006)	0.003 (0.006)	0.004 (0.006)
Averages				
Prop.Medical	-0.021 (0.027)	-0.028 (0.027)	-0.024 (0.027)	-0.022 (0.026)
Interactions				
inter1	-0.017 (0.006)***			
inter2		-0.021 (0.006)***		
inter3			-0.011 (0.006)*	
inter4				-0.016 (0.005)***
Year Dummies				
2010 /11	-0.028 (0.043)	-0.024 (0.043)	-0.027 (0.043)	-0.023 (0.042)
2011 /12	-0.088 (0.046)*	-0.083 (0.046)*	-0.086 (0.046)*	-0.087 (0.046)*
2012 /13	-0.150 (0.047)***	-0.145 (0.047)***	-0.148 (0.047)***	-0.151 (0.047)***
R^2	0.098	0.099	0.097	0.110
Adj. R^2	0.095	0.096	0.094	0.106
Num. obs.	545	545	545	545

This table examines cancellation rates in several clusters, implied by the first regression tree. The skill–mix variables and *Waiting Time* are lagged by one year. Column 1 builds on an indicator variable that takes value one whenever the within–trust means of both *Ideas* and *Feedback* exceed 37 but *Flexible* falls below its median (66.6) and otherwise is zero. The dummy variable in Column 2 is the same except that it takes value one when the within–trust mean of *Flexible* is equal to or exceeds its median (66.6). In Column 3, the binary variable takes value one whenever whenever the within–trust means of *Ideas* exceeds 37, *Decision* lies above 28 and *Job Satisfaction* falls below its median (58.75). The dummy variable is the same in Column 4 except that *Job Satisfaction* is now either equal to or exceeds its median (58.75). The indicator variables are interacted with the proportion of medical staff. All the results are CRE results. The variable of interest in each column appears twice – once in its original form and once as its within–trust mean. With the within–trust mean being included, the coefficient of the interaction term represents the fixed–effects estimate. All the models include year dummies. Robust standard errors are given in parentheses.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

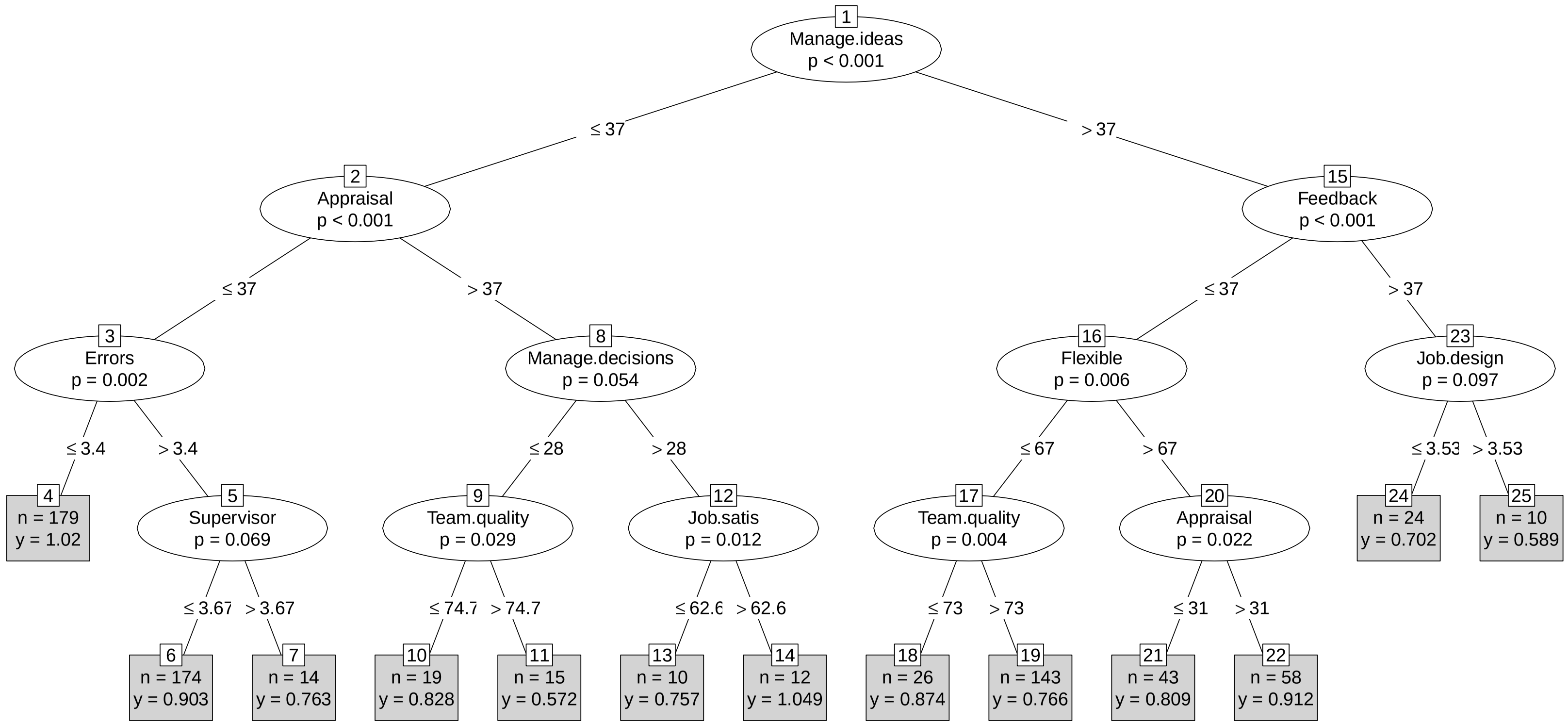
Table 8: Interaction Models: Upper Quartile Thresholds

	(1)	(2)	(3)	(4)
Intercept	0.224 (1.161)	1.017 (1.143)	2.229 (1.804)	1.776 (1.183)
Trust Characteristics				
Teaching	0.035 (0.083)	−0.009 (0.076)	0.012 (0.075)	0.018 (0.077)
Foundation Trust	−0.072 (0.042)*	−0.097 (0.041)**	−0.058 (0.040)	−0.086 (0.041)**
Hospital Beds	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Workforce Skill–Mix				
Prop. Medical	0.001 (0.013)	0.007 (0.012)	−0.010 (0.011)	−0.008 (0.011)
Prop. Nurse	0.008 (0.007)	0.027 (0.009)***	0.007 (0.006)	0.006 (0.006)
Prop. Support	0.007 (0.007)	0.012 (0.007)*	0.008 (0.007)	0.007 (0.006)
Patient Characteristics				
Prop. Emergency	−0.004 (0.004)	−0.004 (0.004)	−0.003 (0.004)	−0.004 (0.004)
Prop. Female	−0.013 (0.010)	−0.016 (0.011)	−0.014 (0.010)	−0.014 (0.010)
Patients aged above 60	−0.003 (0.006)	−0.006 (0.006)	−0.002 (0.006)	−0.001 (0.006)
MFF and Waiting Times				
Inpatient Waiting	0.258 (0.086)***	0.250 (0.084)***	0.232 (0.083)***	0.406 (0.156)***
MFF	0.005 (0.006)	0.006 (0.005)	0.003 (0.005)	0.004 (0.005)
Averages				
Prop. Medical	−0.028 (0.027)			
Prop. Nurse		−0.050 (0.012)***		
Prop. Support			−0.519 (0.363)	
Inpatient Waiting				−0.569 (0.226)**
Interactions				
inter1	−0.023 (0.005)***			
inter2		−0.007 (0.002)***		
inter3			−0.008 (0.003)***	
inter4				−0.082 (0.016)***
Year Dummies				
2010/11	−0.019 (0.042)	−0.013 (0.039)	−0.030 (0.041)	−0.010 (0.037)
2011/12	−0.079 (0.046)*	−0.044 (0.042)	−0.086 (0.044)*	−0.083 (0.045)*
2012/13	−0.141 (0.047)***	−0.109 (0.043)**	−0.151 (0.045)***	−0.153 (0.047)***
R^2	0.107	0.129	0.105	0.108
Adj. R^2	0.103	0.125	0.102	0.105
Num. obs.	545	551	551	545

This table builds on the first unbiased regression tree to construct an indicator variable to capture a possible well–aligned cluster of complementary practices. The skill–mix variables and *Waiting Time* are lagged by one year. The indicator variable takes value one whenever *Ideas*, *Decisions*, *Feedback*, *Job Design*, *Team Quality*, *Job Satisfaction*, *Appraisal*, and *Supervisor* are jointly equal to or exceed their within–trust upper quartile and otherwise is zero. Column (1) interacts the indicator variable with proportion of medical staff. Column (2) interacts the variable with the proportion of nurses. Column (3) interacts it with the proportion of support staff. And Column (4) interacts the variable with *Waiting Time*. The measures of human capital and *Waiting Time* are all lagged by one year. The variable of interest in each column appears twice – once in its original form and once as its within–trust mean. with the within–trust mean being included, the coefficient of the interaction term represents the fixed–effects estimate. All the models include year dummies. Robust standard errors are given in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Figure 3

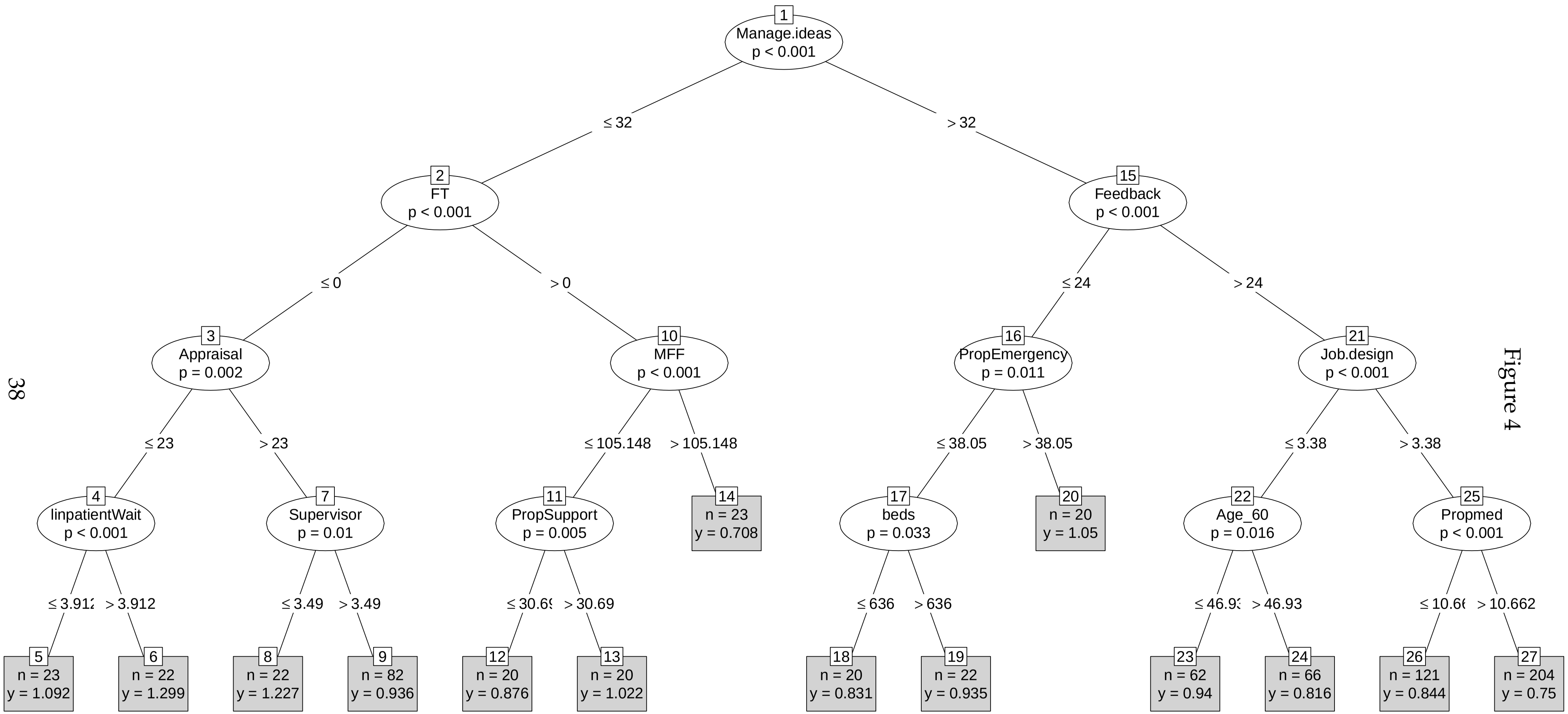
Panel Tree Model I: Management Variables



The unbiased regression tree has access to all the variables listed in the Tree Model I in the text. The maximum depth of the tree is set at four layers for simplicity. The higher a variable appears in the tree, the more predictively important the variable is. *Ideas* appears at the initial node of the tree as predictively the most significant variable and *Appraisal* and *Feedback* as the second most predictively–significant variables. Variables missing from the tree, e.g., *Intention–to–Leave*, are not predictively significant.

Figure 4

Panel Tree Model II: Full Variables



As given in Tree Model II in the text, the unbiased regression tree now also has access to a set of control variables in addition to the variables listed in the base model (Tree Model I). The maximum depth of the tree is set at four layers for simplicity. The higher a variable appears in the tree, the more predictively important the variable is. *Ideas* appears at the initial node of the tree as the most predictively–significant variable. The management variables appear as prominent in the tree. Variables missing from the tree, such as Proportion of nurses (*Prop.Nurse*) and *Decisions*, are not predictively significant.