Economies of Scale and Scope in Hospitals: An Empirical Study of Volume Spillovers

Michael Freeman
INSEAD, Technology and Operations Management Area, 1 Ayer Rajah Avenue, 138676 Singapore
michael.freeman@insead.edu

Nicos Savva
London Business School, Regent's Park, London NW1 4SA, United Kingdom nsavva@london.edu

Stefan Scholtes
Judge Business School, University of Cambridge, Cambridge CB2 1AG, United Kingdom s.scholtes@jbs.cam.ac.uk

General hospitals across the world are becoming larger (i.e. admitting more patients each year) and more complex (i.e. offering a wider range of services to higher acuity patients with more diverse care needs). Although prior work has shown that increased volume is positively associated with patient outcomes, it is less clear how volume affects costs in these complex organizations. This paper investigates this relationship using panel data for 15 specialties comprising both elective and emergency admissions across 157 hospitals in England over a period of ten years. Although we find significant economies of scale for both elective and emergency admissions, we also find evidence of *negative* spillovers across the two admission categories, with increased elective volume at a hospital being associated with an *increase* in the cost of emergency care. Furthermore, for emergency admissions, we find evidence of positive spillovers across specialties – increased emergency activity in one specialty is associated with lower costs of emergency care in other specialties. By contrast, we find no evidence of such spillovers across specialties for elective admissions. Our findings have implications for individual hospital growth strategies and for the regional organization of hospital systems.

Key words: healthcare; productivity; economies of scale; economies of scope; spillovers; econometrics
History: June 8, 2018

1. Introduction

Scale is an important determinant of productivity and a recurrent theme in the operations management and economics literature. Although scale is generally associated with higher productivity (Panzar and Willig 1977), scholars have pointed out that the productivity gains of increased output have to be traded off against the potential productivity losses caused by the increased heterogeneity of that output (Penrose 1959, Schoar 2002). The tension between benefits of scale and potential disbenefits of scope is of particular concern in the hospital industry (Argote 1982, Clark and Huckman 2012). General hospitals provide a large and diverse range of services and use a wide array of technologies and expertise. From both a strategic and operational perspective, this diversity is surprising. At the strategic level, it is at odds with the focus principle (Skinner 1974), and at the process level, it impedes improvement techniques that are based on the reduction-of-variation principle (Hopp and Spearman 2004). Recent studies have discussed the negative impact of the extensive scope of hospital services on service quality measures (e.g. mortality in the hospital context (Kuntz et al. 2017)); however, perhaps due to the lack of sufficiently granular data, the research literature on the productivity (i.e. cost) implications of scale and scope has been less conclusive. This paper uses a novel dataset to provide empirical evidence of the trade-off between scale and scope in the context of the hospital industry.

We focus on two practically relevant sources of heterogeneity in hospital services: (i) *admission-type* heterogeneity (i.e. elective or emergency care) and (ii) *specialty* heterogeneity. Admission-type heterogeneity is the result of collocating the treatment of elective and emergency patients within the same hospital. Elective care is often surgical, ranging from simple day cases (e.g. hernia repairs) and short stays (e.g. joint replacements) to complex, long-stay operations (e.g. open-heart surgery). Elective care is typically planned in advance along a well-defined plan of care. Emergency care has a different dynamic as emergency patients exhibit symptoms that need to be diagnosed and treated under significant time pressure (RCS/DH 2010, AHRQ 2014); there is no a priori treatment plan, and the eventual treatment sequence emerges as a consequence of decisions made "on the spot" as the service progresses. The second source of heterogeneity is related to the medical specialty (also known as service line) that a patient's needs fall under. These are typically organized around specific body parts (e.g. eye, heart), systems (e.g. nervous system, respiratory system), or diseases (e.g. cancer, metabolic diseases) and may share some resources required for patient care (e.g. diagnostic equipment), while other resources are specialty-specific (e.g. specialist physicians). The fact that care of emergency and elective patients of multiple medical specialties are offered within an integrated organization – the general hospital – is as a consequence of the historical evolution of hospitals over the past two centuries (Risse 1999).

From a cost perspective, arguments can be put forward both for positive and negative volume spillover effects (i.e. the effect of a change in volume of one type of activity on another type of activity) that arise from the co-location of multiple medical specialties that treat both elective and emergency patients. On the one hand, pooling spreads fixed costs across a broader customer base (Moore 1959) and can make investing in more productive assets or process structures economical (Argote 2013). Pooling these services may also confer statistical economies of scale by reducing relative arrival variability (Dijk and Sluis 2004), allowing firms to hold less capacity. Finally, pooling provides more opportunities for organizations to learn and accumulate experience (Pisano et al. 2001). On the other hand, it is known that pooling benefits diminish with the degree of dissimilarity between the pooled activities (Joustra et al. 2010, Schilling et al. 2003, Staats and Gino 2012), and these benefits may also be offset by the increase in organizational complexity that comes with mixing customers with different service needs (Argote 1982, Christensen et al. 2009, Kuntz et al. 2017). In summary, whether or not the advantages of pooling counteract the disadvantages of increased organizational complexity in the hospital context is an open question and one that this paper seeks to answer empirically.

Our empirical study is based on annual average cost data of nearly 145 million hospital admissions for over 2,000 conditions treated in 157 acute care hospitals in England over a period of ten years. Since the data is longitudinal and comprised of multiple specialties across multiple hospitals, we estimate the volume effects of interest with within- and between-random-effects multilevel modeling that exploits variation in volume over time and between different specialties/hospitals explicitly (Mundlak 1978, Gelman and Hill 2007). After controlling for changes to asset utilization, we find strong evidence of economies of scale. The more elective patients a hospital treats within a specialty, the lower the cost of these patients (a 10% increase in volume reduces cost per patient by 0.48%). Similarly, if the number of emergency patients in a specialty increases, the cost of these patients decreases (a 10% increase in volume results in a 1.44% reduction in costs). We then focus on volume spillover effects between admission categories as well as spillovers between specialties within admission categories. For electives we find no evidence of spillover effects: An increase in the volume of emergency patients within a focal specialty, or of patients of any admission category across other specialties, has no significant effect on the cost of elective patients in the focal specialty. By contrast, for emergency patients, we find both positive spillover effects – an increase in volume coming from emergency patients in different specialties has a positive impact on emergency productivity (a 10% increase in volume reduces costs by 1.08%) – and negative spillover effects – the volume of elective patients (coming from the focal or other specialties) has a negative effect on emergency productivity (a 10% increase in elective patients within the specialty increases emergency costs by 0.31%, while a 10% increase in elective patients from all other specialties increases emergency costs by 1.34%). These results are robust to alternative model specifications and we can rule out alternative explanations for these findings (e.g. reverse causality, patient selection effects, and endogenous specialty composition).

These findings have important practical implications at both the hospital and regional level. At the hospital level, they suggest that elective care growth strategies – which are often pursued by hospitals to improve overall productivity because elective care has greater standardization potential and, therefore, productivity gains are deemed easier to achieve – may actually lead to a drop in productivity overall because of the unintended negative spillover effect on emergency service productivity. To demonstrate this, we perform a counterfactual analysis based on a large hospital in the metropolitan area of London and show that a 20% increase in hospital admissions across both admission categories leads to a cost saving of 1.3%; however, increasing elective admissions alone by the same number of patients leads to a 2.0% reduction in elective costs but increases emergency costs by 6.7%, leading to a total cost *increase* of 3.3%. Surprisingly, a targeted emergency growth strategy, much less favored by hospital managers due to the complexity of emergency care, is estimated to lead to a cost saving of 7.3% in emergency services without having a significant negative effect on elective care productivity, resulting in a total cost saving of 5.1%.

At the regional level, our results suggest that redistributing hospital services could lead to an aggregate reduction in the cost of providing care. A counterfactual analysis shows that if pairs of hospitals in the London area worked together and redistributed elective specialties so that only one of two hospitals provided any particular service, then the cost of elective treatments could be 3.6% lower without a substantial change in the hospitals' total admissions volumes. Furthermore, our work also presents an additional argument for separating elective patients out of general hospitals. Such patients are better treated in specialized, elective-only treatment centers organized along a single specialty. Physicians and health management researchers have repeatedly called for such reorganization (ASGBI 2007, RCS/DH 2007, Christensen et al. 2009, Bohmer 2009, Hopp and Lovejoy 2012, Monitor 2015), and there is evidence to suggest that this would offer quality benefits across the system (RCS/DH 2007, Kuntz et al. 2017). Our findings complement these studies by providing evidence that such a reorganization would also result in productivity gains. Extending our counterfactual analysis, we estimate, for example, that if London were to operate stand-alone elective treatment centers focused on single specialties only, then elective costs could potentially be reduced by 13.6%. Note, though, that while these counterfactual analyses suggest productivity gains, there may be other reasons – such as quality of care, patient access, or physician preferences – for which such dramatic redesigns may not be implementable.

2. Economies of Scale and Scope in Hospital Care and Related Literature

The empirical literature examining economies of scale in hospitals is quite extensive (see Giancotti et al. (2017) for a recent survey). Although the majority of studies find evidence of the existence of economies of scale, their magnitude and moderating circumstances remain subjects of debate (Aletras 1997, Posnett 2002). From an empirical perspective, identifying the magnitude of scale economies is challenging as estimations may be confounded by unmeasured inter-hospital variation in quality, patient mix and severity, cost accounting and reporting procedures, or the degree of utilization of existing capacity (Dranove 1998, Posnett 2002, Kristensen et al. 2008). The study of scale economies also poses theoretical challenges since economies of scale may arise through several causal mechanisms (Dranove 1998), including the spreading of fixed costs (Moore 1959), learning and innovation (Pisano et al. 2001), and new and better utilization of capacity (Hopp and Lovejoy 2012, Argote 2013). This causal complexity suggests that the degree to which scale affects productivity depends on the organizational level at which the analysis takes place.

Most studies investigate scale economies at either the level of the hospital as a whole (e.g. Marini and Miraldo 2009) or the level of a particular patient condition (e.g. Gaughan et al. 2012). However, the insights into scale effects that can be expected by studying either level in isolation have their limitations (Panzar and Willig 1977). On the one hand, scale at the hospital level is often a consequence of the pooling of heterogeneous services. These studies underestimate the economies achievable through smart pooling of more closely related activities (Dijk and Sluis 2004, Joustra et al. 2010, Vanberkel et al. 2012) to create positive synergies. Studies at the condition level also fail to account for spillover effects among related patient specialties (Schilling et al. 2003, Clark and Huckman 2012). In multi-product firms, these spillovers onto the productivity of one output resulting from a change in the scale of other outputs are referred to as *economies of scope* (Panzar and Willig 1981). Hospital level economies of scale studies thus conflate scale and scope, effectively taking the hospital to be a single-product firm that produces an "average" patient (Kim 1987), while condition level studies disregard the spillover effects onto other services altogether. This study is, to our best knowledge, the first to disentangle these two effects and provide a holistic and managerially relevant account of economies of scale and scope in hospitals.

To study scale and scope economies, we examine the cost implications of general hospitals treating together patients who may differ along two dimensions: their admission category (elective or emergency admission) and the medical specialty associated with their condition (e.g. cardiac, respiratory, etc.). Past literature suggests that scale improves productivity as it allows to spread fixed cost more widely (Moore 1959), makes an investment in more productive assets or process more attractive (Argote 2013), reduces variability (Dijk and Sluis 2004), and accommodates faster learning through accumulate experience (Pisano et al. 2001). Thus, past literature can be used to hypothesize that treating more patients of the same admission category within a specialty (e.g. elective patients with a cardiac condition) should allow hospitals to deliver care at a lower cost for these patients. Nevertheless, the existence of productivity spillover effects (positive or negative) to other specialties or other admission categories is less clear. In particular, these spillover effects may depend on the degree to which the other activities are related to the focal activity. When inputs are shared or utilized jointly by related activities, synergistic economies can be realized resulting in reduced costs of production across activities (Panzar and Willig 1977, Hill and Hoskisson 1987). Porter (1985) distinguishes between two possible sources of such synergies: those arising from tangible interrelations between activities – resulting from, e.g., the sharing of raw materials, technology, and production processes – and those arising from intangible interrelations – resulting from, e.g., opportunities to apply learning from one situation to another. Thus the more related the activities, the more advantages there are to be gained from providing these activities alongside each other at higher volumes. Overall, these mechanisms create positive productivity spillovers from increased scale of one activity to other activities, since this enables greater exploitation of synergies. At the same time, the advantages should be stronger when increasing the volume of patients of the same admission category and specialty than increasing the volume of patients of another admission category and/or another specialty, as the tangible and intangible interrelations are also stronger.

Although pooling patients from different specialties and admission categories will create synergies, Christensen et al. (2009) point out that there are operational tensions between two fundamentally different operational processes within hospitals which are likely to cause counteracting negative spillover effects. On the one hand, hospitals treat patients with poorly specified and often urgent needs. The quality and efficiency of their care depend on the speed and accuracy of the search process for the root cause and the most appropriate treatment. This process is often highly variable, and so benefits from effective knowledge exchange between different specialties. Christensen et al. (2009) call this process "intuitive medicine". On the other hand, hospitals treat patients with well-diagnosed conditions and a clear treatment plan. The service for these patients is not as time-critical. The treating physician typically assesses them in an outpatient office before admission to the hospital and their symptoms are well-diagnosed before a hospital appointment is made to carry out a clearly defined procedure. To be effective and efficient, these care processes should leave no room for trial and error and deliver predictable outcomes consistently. Christensen et al. (2009) call these processes "precision medicine." While pure intuitive medicine and pure precision medicine are extreme cases and much of the day-to-day operation in a hospital falls somewhere in-between, emergency patients rely more on the practice of intuitive medicine, while elective patients benefit more from precision medicine.

Because the operational character of delivering intuitive and precision medicine are not always aligned, the effectiveness and efficiency of these services may deteriorate when they co-exist in the same organization (Christensen et al. 2009). More specifically, tensions arise because there are substantial differences in the optimal configuration of hospital assets (e.g. operating theatres, patient wards, diagnostic labs) and patient pathways of care (e.g. clinical investigations and diagnosis, admission, treatment, discharge) for intuitive (emergency) and precision (elective) medicine. For example, to cater for the former the hospital may require specialist physicians to respond to Emergency Department requests to review patients within a specific time-frame (at the expense of their elective activities), interdisciplinary collaboration to accurately diagnose conditions and devise appropriate treatment, and may instigate preemptive priority to operating theater schedules and beds leading to elective cancelations (Nasr et al. 2004). To cater for the latter, a hospital may ring-fence wards and operating theatres for elective patients only, which, given the more predictable nature of elective care, can operate at higher utilization than their emergency or mixed-use counterparts (Kjekshus and Hagen 2005). Naturally, as the volume of one of the two admission categories increases, the optimization of hospital assets and pathways of care will be skewed more towards it, generating positive productivity gains for such patients at the expense of the other admission category that will have fewer dedicated resources or will have to be treated by assets and pathways optimized for something else, generating a negative spillover effect. This negative spillover effect may counteract the positive spillover effects caused by tangible and intangible interrelations. As a consequence, while one might be able to hypothesize positive spillover effects between specialties within admission categories, it is not possible to hypothesize the sign of the spillover effects across admission categories, and it becomes necessary to estimate the direction of the aggregate effect empirically.[1]

The answer to this question, which has important practical implications, cannot be deduced by examining the empirical evidence from other industries. Although evidence from other industries shows that economies of scale are, for the most part, pervasive (Junius 1997), there is conflicting evidence as to the extent and direction of scope effects. Benefits have been demonstrated to exist in contexts such as drug R&D (Henderson and Cockburn 1996) and advertising (Silk and Berndt 1993), while diseconomies have been found in others such as transportation (Rawley and Simcoe 2010) and automobile assembly (Fisher and Ittner 1999). In industries such as manufacturing

[1] We note that prior work has investigated spillover effects and economies of scope in hospital care. However, these studies have had significant data limitations. The majority use hospital-level annualized costs and can only distinguish between scope effects arising from the co-production of hospital services at a high level of aggregation, e.g. between inpatient, outpatient, and ambulatory services as opposed to individual medical specialties (Preyra and Pink 2006, Carey et al. 2015). Furthermore, these studies are often limited to small single-year samples, raising concerns about unobserved heterogeneity and making it challenging to establish causality (Monitor 2012, Gaynor et al. 2015). A summary of the findings of these studies can be found in Section EC.10 of the supplementary material.

(Kekre and Srinivasan 1990, Schoar 2002), airlines (Gimeno and Woo 1999, Tsikriktsis 2007) and education (Sav 2004) the evidence is often conflicting and may depend on the level of analysis. Given that prior work suggests that scope effects may be context specific, coupled with the fact that the hospital sector has a number of idiosyncratic differences to other industries, the measurement of productivity pullovers requires an empirical approach.

A stream of literature complementary to studies of economies of scale investigates how volume and focus affect the quality of patient care in hospitals. In their studies on performance in cardiothoracic surgery, Pisano et al. (2001) show that as surgeons perform more procedures they accumulate experience and become faster, while Huckman and Pisano (2006) find that this is also associated with a reduction in mortality, although this effect is firm-specific, and KC and Staats (2012) identify that the reduction in mortality associated with learning is greater if surgeons perform a larger volume of focal tasks rather than similar but related tasks (see also Ramdas et al. 2017). The degree to which task similarity moderates the volume–outcome relationship has also been studied in the focus literature. Clark and Huckman (2012) find that cardiovascular patients experience better clinical outcomes when a hospital specializes in cardiovascular care but also that there are positive spillovers for these patients if the hospital provides complementary ancillary services as well. This finding is complemented by a number of studies outside the healthcare context, with Schilling et al. (2003) showing that there are learning benefits associated with performing both repeated and related tasks but not with unrelated activities (see also Boh et al. 2007, Narayanan et al. 2009, Staats and Gino 2012). Our work differs from these studies in its focus on the debate on productivity (i.e. the cost of providing care) rather than quality (e.g. patient mortality) as well as in the highly relevant investigation of the productivity spillover effects of volume between different specialties and admission categories.

Finally, our work relates to a large and growing stream of empirical operations management literature that examines the impact of organizational workload on operational performance and patient outcomes in hospital care. Recent examples include KC and Terwiesch (2009), Powell et al. (2012), KC and Terwiesch (2012), Green et al. (2013), Kuntz et al. (2014), Kim et al. (2014), Chan et al. (2017), Freeman et al. (2017a,b), and Batt and Terwiesch (2017), among others. In contrast to this literature, which exploits short-term temporal variation in workload, our work focuses on the more long-term impact of volume on hospital costs. Our identification strategy exploits variation across hospitals and specialties after controlling for changes in utilization over time.

3. Data, Variable Definitions, and Econometric Models

The primary data set for this study consists of annual costing and inpatient activity data for the ten financial years from 2006/07 to 2015/16 for all acute hospital trusts operated by the National

Health Service (NHS) in England. Acute NHS hospital trusts provide secondary and tertiary care in facilities that range from small district hospitals to large teaching hospitals. Services include EDs, inpatient and outpatient medicine and surgery, and specialist medical services. We focus our attention on admitted patient care and exclude outpatient activity and ED visits that do not result in hospital admission. In total, our data comprises aggregate annual information for nearly 145 million patient admissions to 157 acute hospital trusts. Since a number of trusts were merged during the observation period, whenever a trust merges with another we treat the new organization as a distinct entity, increasing the effective number of trusts from 157 to 169.

For regulatory purposes, each hospital trust is mandated to complete an annual return of so-called reference costs, reporting the trust's activity for each patient condition treated over the preceding year. Patient conditions are defined using so-called healthcare resource groups (HRGs), which are the UK equivalent of the diagnosis-related groups (DRGs) used by Medicare in the US. HRGs are designed so that patients within an HRG are clinically similar and require a relatively homogeneous bundle of resources for their treatment (Fetter 1991). Each patient admission is assigned to a unique HRG using an automated process based on information provided in the discharge notes, including standardized ICD-10 medical diagnosis codes, OPCS procedure codes, and contextual information such as patient age, gender and the existence of any complications or comorbidities (see e.g. DH 2013). The costs incurred by a hospital each year are allocated to specific HRGs, with each hospital reporting the average cost of treating patients within each HRG, the average length of stay (LOS) of these patients, and the volume of patients treated from each HRG. The primary data set is comprised of just under 10.4 million of these HRG-level submissions.

These cost submissions are used by the UK Department of Health to determine the price (also known as the "tariff") to be paid to hospitals for each discharged patient in an HRG in the following financial year. While the specifics are complex, the main principle is to reimburse hospitals per patient at a rate that is close to the national average cost of providing treatment for the specific HRG to which each patient is assigned. The intention behind this benchmarking approach is to generate cost reduction incentives (see Shleifer 1985, Savva et al. 2018). Since the reported costs are critical for hospital reimbursement, it is of paramount importance that they are reliable and comparable across hospitals. To ensure that this is the case, hospitals are issued with extensive guidelines on how to allocate direct, indirect, and overhead costs to different HRGs (e.g. HFMA 2016) and the UK Department of Health commissions regular independent audits. In 2010, halfway through our observation period, the UK Audit Commission, a statutory corporation that performs regular audits of public bodies in the UK, performed a comprehensive audit of the data accuracy of seven years of NHS reference cost submissions (UKAC 2011). The report concluded that "most trusts' reference costs submissions were accurate in total." Nevertheless, the report also noted that "the accuracy of individual unit costs varied and, in some cases, was poor." We address this point in our definition of specialties.

Specialty Categories. Although each individual HRG can be thought of as a distinct specialty, we have chosen to define specialties at a coarser level for two reasons. First, HRG codes are updated annually and have become more granular over time; the number of HRG codes in our data increases every year, from 1,149 in 2006/07 to 2,440 in 2015/16, leading to a total of 4,749 unique HRG codes in our data. To account for this change in coding over time, we are able to map these 4,749 codes to a set of 496 time-invariant HRG roots – using a publicly available data source intended for this purpose (HSCIC 2015) – which group similar HRGs together. Each HRG root then falls within one of 21 HRG chapters, which we subset to 16 clinically meaningful core HRG chapters that correspond to the major body systems, e.g. nervous or respiratory system, or to particular medical specialties, e.g. obstetrics or cardiac conditions.[2] Although two identical patients in different years may be assigned different HRG codes or, to a lesser extent, different HRG roots, it is unlikely that they would be allocated to different HRG chapters. The HRG chapters, therefore, provide time-consistent clusters of patients with related conditions, which we define as medical specialties for the purpose of this study.

The second reason for choosing this higher level of aggregation has to do with concerns about the reliability of cost allocations at the individual HRG level. Cost allocation conventions for specific HRG codes *within* HRG chapters can vary significantly between hospitals, but any such deviations within chapters average out when aggregated to the chapter level. This results in considerably more consistent cost allocations at the HRG chapter level. This was confirmed by a former director of costing at the UK Healthcare Financial Management Association, the main advisory body for the financial governance of hospitals in the UK. We note that a similar aggregation approach to that described above has been adopted in related empirical research (e.g. Greenwald et al. 2006, Clark 2012, Clark and Huckman 2012). A list of the specialties (i.e. HRG chapters) included in this study appears in the caption of Figure 2.

[2] Of the five HRG chapters that we drop, four correspond to so-called "unbundled" activities. These are additional, exceptional, high-cost or non-routine elements of care that are reported and reimbursed separately from the core HRG. Examples include radiotherapy, diagnostic imaging, rehabilitation, renal dialysis, chemotherapy, and high-cost drugs. The fifth chapter dropped is used in rare cases where patients cannot be assigned to an HRG, which occurs in only 0.06% of cases. In total these unbundled activities constitute only 6.05% of total costs. In our analysis we include a control variable to account for this excluded activity – see Section 3.4.

To further alleviate concerns about the reliability of cost accounting, we corroborate the results of the costing analysis with a length-of-stay (LOS) analysis; LOS does not suffer from accounting errors (as patient admission and discharge dates are easy to capture) and is highly correlated with hospital costs.

Admission Categories. Every hospital reports for each HRG the costs, volume, and LOS for three patient admission categories: (1) day cases, (2) elective inpatients, and (3) emergency (non-elective) inpatients. In contrast to emergency admissions, elective inpatient and day-patient admissions are scheduled in advance, with the former including at least one overnight stay in a hospital bed. When the national tariff for an HRG is calculated the standard approach is to treat day cases and elective inpatients as substitutable and to reimburse at the same rate. We follow this approach and merge day cases and elective inpatients, leaving two admission categories: electives (El) and emergencies (Em).

Note that elective and emergency patients may be assigned to the same HRG code but, importantly for our analysis, the costs, LOS, and activity data are reported separately for each admission category. One complication is that, due to a coding convention, all obstetric activity is recorded as emergency/unplanned (and insufficient information is available to manually separate this out into elective versus emergency activity). Therefore, we have removed the specialty for obstetric services from the sample. Since obstetrics typically operates as a stand-alone service within a hospital this is unlikely to have much bearing on the results. However, as it accounts for 9.1% of costs amongst the core HRGs we will include in our models a control variable to account for its removal – see Section 3.4.

Data Hierarchy and Unit of Analysis. Within each admission category (emergency or elective), each observation belongs to two (non-nested) levels: the specialty and the hospital trust. Time is a third level. The data set contains 21,510 specialty-trust-years across 15 medical specialties observed longitudinally over 1,434 trust–years. After removing three specialty–trust–years where no data in the multiple trauma specialty was observed, we obtain 21,507 specialty–trust–years for the analysis of emergency admissions. For elective admissions, we drop the multiple trauma specialty, for which all patients are emergency admissions, and 19 specialty–trust–years for which no patients were admitted in that specialty–trust–year, resulting in 20,057 observations for the analysis of elective admissions.

3.1. Dependent Variables

The main dependent variables in this study are the average costs per patient for (a) emergency and (b) elective hospital admissions. As discussed above, we complement this analysis with an additional measure, the average LOS per patient for the two admission categories. For the purposes of our study we adjust the average cost and LOS per admission-type–specialty–trust–year to account for (i) cost variation between hospitals due to regional factors, (ii) cost and LOS variation within a specialty between hospitals, due to differences in the case-mix within the specialty, and (iii) heterogeneity in the cost and LOS distribution between specialties and over time. Our approach is very similar to that used by the UK government to calculate hospital-level reference cost indices for comparing the relative efficiency of hospitals (see e.g. DH 2016, Chp. 4), except that we adjust costs and LOS at the level of the specialty instead of a hospital, and also introduce step (ii). We provide more details on these adjustments below.

Regional Cost Adjustment. We account for regional differences as costs may vary due to local factors outside the hospital trusts' control, e.g. regional variation in the cost of wages, land, and buildings. We do this by adjusting the reported average costs per patient using a government-produced market forces factor (MFF) designed for this purpose (Monitor 2013). The MFF, given by m_{th}, is a scalar unique to each hospital trust h in each year t that is used to weight its costs based on the level of unavoidable spending faced relative to other trusts. Specifically, the regionally adjusted cost for a patient of admission category $p \in \{El, Em\}$ assigned to HRG code c in hospital trust h and year t is equal to $\text{cost}_{thcp} = \frac{\text{cost}'_{thcp}}{m_{th}}$, where cost'_{thcp} are the costs reported in the data.

Case-mix Adjustment. As explained earlier in this section, we aggregate data from the HRG level to the specialty level (HRG chapter). Differences in the average regionally adjusted cost per specialty patient between two hospitals could therefore be due to a different HRG case-mix within the specialty. Take, for example, a specialty with two HRGs X and Y and suppose costs of HRG X are lower than those of HRG Y, independently of the hospital that treats these patients. If Hospital A has 30% of its specialty patients in HRG X and 70% in Y, while Hospital B has 10% in X and 90% in Y, then this case-mix difference will cause Hospital A's average cost per patient in the specialty to be lower than Hospital B's, simply because it treats relatively more patients in the cheaper HRG X. To adjust for this case-mix effect, we *do not* calculate a hospital's average cost per specialty patient based on the individual hospital's relative volumes of HRGs in the specialty (i.e. $30\%\text{Cost}_{\text{XA}} + 70\%\text{Cost}_{\text{YA}}$ for Hospital A and $10\%\text{Cost}_{\text{XB}} + 90\%\text{Cost}_{\text{YB}}$ for Hospital B) but instead fix the same relative volumes accross all hospitals (e.g. choose relative volumes, say 20% and 80%, and calculate the costs of a specialty patient as $20\%\text{Cost}_{\text{XA}} + 80\%\text{Cost}_{\text{YA}}$ for Hospital A and $20\%\text{Cost}_{\text{XB}} + 80\%\text{Cost}_{\text{YB}}$ for Hospital B). This amounts to projecting the average cost per specialty patient in the hospital, conditional on the same case-mix for all hospitals. We choose this fixed case-mix based on the set of 116 reference trusts, T_r, (74% of all trusts in the data) that we observe throughout the entire observation period and that have not been involved in a hospital merger. We aggregate their HRG volumes, and calculate the relative volumes of individual HRGs in a specialty in this aggregated *reference trust*. We perform this case-mix adjustment separately for each observation year and admission category and adjust LOS analogously.

Formally, let C_{tp} be the set of HRGs c in specialty C observed in year t for patients of admission category p. Then the weight (i.e. relative volume) assigned to a particular HRG $c \in C_{tp}$ is equal to $\alpha_{tcp} = \frac{n_{tcp}}{\sum_{c \in C_{tp}} n_{tcp}}$, where n_{tcp} is the total number of patients across all reference trusts $h \in T_r$ of admission category p with HRG c in year t. Then hospital trust h's average cost, $\mathbf{Cost}_{thCp}$, for patients of admission category $p \in \{El, Em\}$ in specialty C and year t is calculated as

$$\mathbf{Cost}_{thCp} = \sum_{c \in C_{thp}} \alpha_{tcp} \mathrm{cost}_{thcp}\,, \tag{1}$$

where $C_{thp} \subseteq C_{tp}$ is the subset of HRGs c in specialty C for patients of admission category p that are observed in trust h in year t. We perform a similar weighting procedure to calculate the case-mix-adjusted average LOS.

Cost Standardization. After case-mix adjusting, costs within a specialty in a given year can be compared across hospitals. However, costs may still vary across specialties (e.g. between cardiac conditions and conditions related to the eyes) and over time (due e.g. to macroeconomic factors, such as inflation, or changes in guidance or regulation that are common to all hospital trusts and that may render specific specialties more (or less) costly). We could account for this by including e.g. specialty and year fixed effects in the econometric models, which would act to de-mean the case-mix adjusted average costs. However, if the variance of costs differs across one or more of the three levels of our panel then the errors (residuals) will be heteroskedastic even after de-meaning – a violation of the IID assumption. The left-hand column of Figure 1 shows that heteroskedasticity of costs across specialties exists even after de-meaning. To reduce heteroskedasticity, we divide $\mathbf{Cost}_{thCp}$ with the corresponding case-mix adjusted expected cost, calculated from a set of comparator trusts. The comparator trusts, T_h, for each hospital trust h is the set of 116 reference trusts described earlier, excluding hospital h if $h \in T_r$ (to ensure that the relationship between costs and expected costs is not endogenous), i.e. $T_h = T_r \setminus \{h\}$.

Formally, we define the expected cost of an HRG c to be equal to

$$\overline{\mathrm{cost}}_{thcp} = \frac{\sum_{h \in T_h} n_{thcp} \mathrm{cost}_{thcp}}{\sum_{h \in T_h} n_{thcp}}\,, \tag{2}$$

where n_{thcp} is the number of patients of admission category $p \in \{El, Em\}$ assigned to HRG code c in hospital trust h and year t. The expected cost of treating an average patient from specialty

C at hospital h is then be calculated by replacing cost_{thcp} in Equation (1) with $\overline{\text{cost}}_{thcp}$, giving $\overline{\mathbf{Cost}}_{thCp}$. Taking the ratio of $\mathbf{Cost}_{thCp}$ to $\overline{\mathbf{Cost}}_{thCp}$ gives the case-mix adjusted and normalized costs. A similar adjustment is made for LOS.

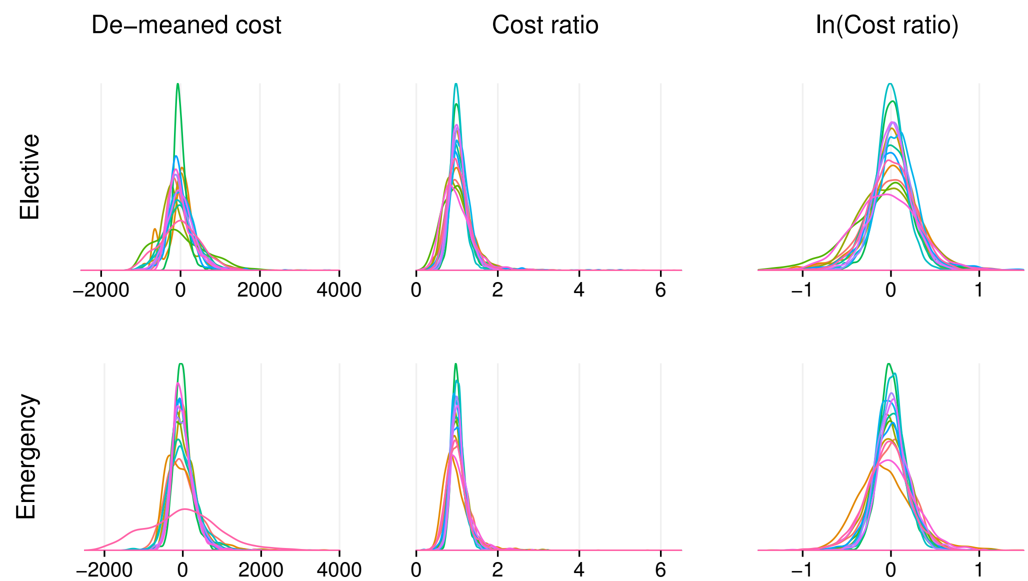
To see how this works, suppose that inflation causes costs to increase by 3% in all hospitals. Then expected costs would then also increase by 3%, and so taking the ratio would remove the inflationary effect. Further, if costs are, say, 20% higher in specialty A than in specialty B, then the expected costs will also be 20% higher in specialty A. As a further advantage, observe that there is no guarantee that a hospital trust will see patients from every HRG c from specialty C in every financial year. This means that while $\sum_{c\in C_{tp}} \alpha_{tcp} = 1$, it might be the case that $\sum_{c\in C_{thp}} \alpha_{tcp} < 1$, since one or more $c \in C_{tp}$ may not be in C_{thp}. In Equation (1) this would have the effect of reducing $\mathbf{Cost}_{thCp}$, artificially deflating our cost measure and making across-hospital comparisons problematic. Notice, though, that $\overline{\mathbf{Cost}}_{thCp}$ will be reduced also, since it is calculated over the same set of HRGs $c \in C_{thp}$ as is $\mathbf{Cost}_{thCp}$. As a result, taking the cost to expected cost ratio will adjust for any unobserved HRGs and so ensures that costs remain comparable across hospitals (effectively by assuming that those unobserved HRGs would have been above or below expected cost to the same extent as all of the HRGs that are observed).

In summary, differentiating between elective and emergency admissions, we obtain the four dependent variables: $CostEl$ and $CostEm$, the regionally, case-mix-, and standardized average costs per elective and emergency patient, respectively, and $LOSEl$ and $LOSEm$, the average case-mix- and standardized LOS for elective and emergency patients, respectively. An example demonstrating further the construction of the dependent variables can be found in Section EC.8 of the supplementary material. The distribution of the cost variables for each specialty (and the distribution of their logarithm) are shown in the middle (right) column of Figure 1. Any differences in costs or LOS between hospital trusts and specialties that are not accounted for by this adjustment method will be captured through the control structure of the econometric models.

3.2. Independent Variables

To investigate economies of scale and scope we use four measures of volume: the volume of (i) elective, $nElS$, and (ii) emergency, $nEmS$, activity within a specialty (the focal specialty) and the volume of (iii) elective, $nElH$, and (iv) emergency, $nEmH$, activity from all specialties *other than* the focal specialty. Volume refers to the total number of patient admissions per annum. Throughout, we log transform all volume measures to reduce heterogeneity across specialties, skewness, and the influence of outliers.

Figure 1 Distribution of cost by specialty: De-meaned average costs by specialty (left), average cost ratios (middle) and the natural logarithm of the ratios (right), for elective (top) and emergency (bottom) admissions.



3.3. Econometric Specification

To simplify the hierarchical structure of the data we present the main analysis using two distinct panels: one for emergency and one for elective patients.[3] Each observation within a panel belongs to three (non-nested) levels: specialty, hospital trust, year. In this section we present the models for the costs of elective patients; the equivalent models for emergency costs or for LOS can be formulated by replacing the dependent variables accordingly.

The econometric analysis deploys the Mundlak (1978) within–between formulation in the multilevel modeling (MLM) literature (Certo et al. 2017). Although within–between MLMs are frequently used in other fields, they are less common in the operations management literature despite their numerous advantages (Bell and Jones 2015). Estimating a within-between MLM requires that the continuous covariates are decomposed into (1) the cross-sectional (i.e. between-hospital) variation, and (2) the longitudinal (i.e. within-hospital) variation. The measures of cross-sectional volume variation captures differences in the aggregate sizes of the specialties at the different hospitals. These are given by calculating the average of each of the four volume measures for each hospital–admission-type–specialty over the observation period. For example, if $nElS_{thC}$ gives the number of elective patients in specialty C in hospital trust h in year t, then the corresponding cross-sectional volume after taking the natural logarithm is given by $nElS^{CS}_{hC} = \sum_t \frac{ln(nElS_{thC})}{n_h}$, where n_h

[3] We can combine the two panels and estimate the results jointly, which results in quantitatively and qualitatively similar findings – see Section EC.7 of the supplementary material for details.

is the number of years that hospital trust h is observed in the data set. Using this approach we generate the four cross-sectional volume measures, $nElS^{CS}$, $nEmS^{CS}$, $nElH^{CS}$ and $nEmH^{CS}$. The measures of longitudinal volume variation, on the other hand, capture the effect of a (usually small) change in volume within the same hospital over time. These are calculated by subtracting the cross-sectional volume from the natural logarithm of the raw volume observed in a given year, e.g. $nElS^{LT}_{thC} = ln(nElS_{thC}) - nElS^{CS}_{hC}$, giving the four longitudinal volume measures $nElS^{LT}$, $nEmS^{LT}$, $nElH^{LT}$ and $nEmH^{LT}$. Summary statistics for costs, LOS, and cross-sectional and longitudinal specialty and hospital volume for both the elective and emergency patient segments appear in Table 1.

These two types of volume measure different effects. The cross-sectional volumes capture the approximate scale of the focal and non-focal specialties at each hospital, as well as how this is split between elective and emergency activity. This is likely to capture the different hospital asset configuration and patient pathway optimization associated with more volume and allows us to answer the question: are costs on average lower in higher volume hospitals than in lower volume hospitals? The longitudinal volume measures allow us to identify how costs respond to the small and gradual changes in the volume of patients treated at the same hospital–specialty at different points in time, assuming capacity to be fixed.[4] (Indeed, as can be seen in Table 1, the cross-sectional variability of volume, as measured by the standard deviation, is 2.6–7.4 higher than the longitudinal variability.) This measure addresses the question: how sensitive are costs to small perturbations in the volume of patients that they treat over time? In other words, one can think of the cross-sectional effect as the *hospital design effect* (controlling for variation in utilization over time), while the longitudinal effect captures the *asset utilization effect* (controlling for hospital "design"). It is the former that captures the main scale and scope effects of interest in this study, while the latter serves to measure and control for how costs respond to changes in asset utilization at different points in time. For more on the distinction between these two types of effect and discussion of why the cross-section effect is more relevant to our study see Section EC.9 of the supplementary material. Our econometric approach will therefore focus on how to identify the impact of the four cross-sectional (between-hospital) volume measures on cost.

[4] We show that the assumption of fixed capacity can be relaxed in our robustness tests in Section EC.6.5 of the supplementary material, but note also that the asset configuration of UK hospitals is likely to have remained relatively stable during the observation period: In the wake of the 2008 global financial crisis, the national government decided essentially to freeze the NHS budget in real terms, despite continuously increasing demand pressure (NAO 2011, HMT 2015, NT 2016), making it difficult for hospitals to find the capital to invest in significant changes to asset structures.

Table 1 **Descriptive statistics and correlation table**

		Descriptive statistics				Correlation table							
	Variable	Mean	SD	Min	Max	(1)	(2)	(3)	(4)				
(1) Elect. cost / exp. cost	$CostEl$	1.05	0.51	0.11	54.40		0.13^{***}	0.22^{***}	0.00				
(2) Emerg. cost / exp. cost	$CostEm$	1.02	0.24	0.07	5.67			0.07^{***}	0.41^{***}				
(3) Elect. LOS / exp. LOS	$LOSEl$	1.04	0.21	0.42	12.98				0.16^{***}				
(4) Emerg. LOS / exp. LOS	$LOSEm$	1.03	0.18	0.12	4.77								
	Variable	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(5) LT ln(elect. service vol.)	$nElS^{LT}$	-0.00	0.30	-4.00	3.28	-0.09^{***}	-0.01	-0.12^{***}	0.00		0.17^{***}	0.26^{***}	0.13^{***}
(6) LT ln(emerg. service vol.)	$nEmS^{LT}$	0.00	0.23	-2.04	1.93	-0.00	-0.15^{***}	-0.04^{***}	-0.13^{***}			0.33^{***}	0.67^{***}
(7) LT ln(elect. hospital vol.)	$nElH^{LT}$	0.00	0.11	-0.83	0.62	-0.05^{***}	-0.00	-0.09^{***}	0.03^{***}				0.47^{***}
(8) LT ln(emerg. hospital vol.)	$nEmH^{LT}$	-0.00	0.17	-1.16	0.81	-0.01	-0.12^{***}	-0.06^{***}	-0.09^{***}				
	Variable	Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(9)	(10)	(11)	(12)
(9) CS ln(elect. service vol.)	$nElS^{CS}$	6.76	2.23	0.00	10.24	0.00	0.07^{***}	-0.11^{***}	-0.05^{***}		0.46^{***}	0.17^{***}	0.19^{***}
(10) CS ln(emerg. service vol.)	$nEmS^{CS}$	7.37	1.26	2.63	10.10	-0.01	0.04^{***}	0.06^{***}	-0.08^{***}			0.29^{***}	0.27^{***}
(11) CS ln(elect. hospital vol.)	$nElH^{CS}$	10.39	0.50	8.68	11.61	0.04^{***}	0.13^{***}	0.02^{*}	-0.03^{***}				0.85^{***}
(12) CS ln(emerg. hospital vol.)	$nEmH^{CS}$	10.53	0.45	9.05	11.74	0.03^{***}	-0.00	0.00	-0.10^{***}				

Notes: LT denotes the longitudinal volume effects; CS denotes the cross-sectional volume effects; $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

More specifically, the econometric model we estimate takes the following form:

$$\begin{aligned} ln(CostEl_i) = \alpha_{(thC)[i]} &+ \beta_1^{LT} nElH_i^{LT} + \beta_2^{LT} nElS_i^{LT} + \beta_3^{LT} nEmH_i^{LT} + \beta_4^{LT} nEmS_i^{LT} \\ &+ \beta_1^{CS} nElH_i^{CS} + \beta_2^{CS} nElS_i^{CS} + \beta_3^{CS} nEmH_i^{CS} + \beta_4^{CS} nEmS_i^{CS} + \epsilon_i \,, \end{aligned} \tag{3}$$

where the (random) intercept is given by

$$\alpha_{(thC)[i]} = \mathbf{bX} + \beta^t P_{(t)[i]} + \beta^C P_{(C)[i]} + \alpha_{(h)[i]} + \alpha_{(th)[i]} + \alpha_{(tC)[i]} + \alpha_{(hC)[i]} \,. \tag{4}$$

Using the notation recommended in Gelman and Hill (2007), the index $(thC)[i]$ denotes the time, t, hospital trust, h, and specialty, C, corresponding to observation i, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the idiosyncratic error term. The variables P_t and P_C are time and specialty FEs, respectively and the vector $\mathbf{X}$ represents controls which we will discuss in Section 3.4 below.

We make two observations. First, the specification of the random intercept, $\alpha_{(thC)[i]}$, makes this model more flexible than traditional fixed-effect (FE) regression techniques. The terms $\alpha_{(x)[i]}$, where $(x)[i]$ takes values $(h)[i]$, $(th)[i]$, $(tC)[i]$, and $(hC)[i]$, denote the hospital trust, trust–year, specialty–year and specialty–trust random effects (REs), respectively, which are all assumed to follow a Normal random variables with a standard deviation to be estimated.[5] Second, formulating the model as a within-between MLM as opposed to a simple RE overcomes one of the drawbacks of the RE model, the assumption that random intercepts are not correlated with the independent variables (e.g. the volume). If this assumption is violated (e.g. if there are unobservable factors such as

[5] We could also have estimated the time and specialty FEs as REs in Equation (4), since the number of categories (10 years and 14 specialties) and large amount of data per category makes the RE estimation qualitatively similar to that for FE (Gelman and Hill 2007). The results are indeed similar if we estimate these as REs instead.

"management quality" that make a hospital more likely to have both high cost realization and high volume), then the estimated coefficients would suffer from heterogeneity bias and the errors would be unreliable (Hsiao 2015). The MLM model offers an elegant solution to this problem by including the average of the dependent variables explicitly in the model (Mundlak 1978). Furthermore, this formulation also has a number of other advantages, including the fact that correct standard errors are automatically estimated without resorting to error clustering (Bell and Jones 2015), and that this model allows us to also add higher-level variables (i.e. variables that would have otherwise been collinear with fixed effects in FE models) as controls. This can help to reduce the (unexplained) variability in the random error. In the next section we introduce a number of such controls.

3.4. Controls

There are various factors that confound the effect of volume on costs. By exploiting the panel structure and through the inclusion of the fixed- and random-effects, the multilevel control structure adjusts for many of these. For example, factors specific to a hospital or a specific specialty within a hospital (e.g. local competition, complexity of the patient pool, patient demographic and socioeconomic status) or those specific to a hospital but that might change over time (e.g. management, facilities and equipment) are already accounted for. However, where possible, we identify additional controls to include in our models and discuss them below.

Some hospitals may elect to provide a full range of services within a particular specialty, while others may choose to concentrate on treating particular conditions. Since this may affect the cost structure, we include four controls (two for electives, two for emergencies) that measure the range of conditions treated and the degree of concentration. The first two controls measure the proportion of elective (emergency) services offered within the focal specialty in a given hospital in a particular year. This is calculated by summing over the weights α_{tcp} defined in Section 3.1, and is equal to $\mathbf{Prop}_{thCp} = \sum_{c \in C_{thp}} \alpha_{tcp} \leq 1$. When $\mathbf{Prop}_{thCp} = 1$ then the hospital provides treatment across the full range of conditions, and the closer to 0 the more narrow the range of conditions within a specialty that a hospital offers. The second two controls capture the extent to which a hospital's elective (emergency) activity within a specialty is concentrated within a small (or spread across a large) number of HRGs. This concentration measure is based on the Herfindahl-Hirschman Index. Specifically, if a_{thcp} is the proportion of elective (emergency) activity concentrated in HRG c within specialty C at trust h in year t, then $\mathbf{Conc}_{thCp} = \sum_{c \in C_{thp}} a^2_{thcp}$ is a measure of the overall concentration of activity within specialty C. Both of these controls are interacted with the specialty fixed effect, P_C, to capture possible heterogeneous effects across specialties.

One point made by extant literature is that a change in volume in one dimension, with volume held constant in all other dimensions, will also change the "focus" of the hospital (e.g. McDermott et al. 2011). So as not to confound the effect of volume spillovers with that of focus, we introduce another two variables (one for electives, one for emergencies) based on the Herfindahl-Hirschman Index. These variables serve to capture the degree to which hospitals are differentiated in terms of their service mix across specialties. This is equal to the sum of squared shares (hospital-specific, not across all hospitals) of elective (emergency) volume for each for each of the specialties, and is given by $\mathbf{Conc}_{thp}$. This is a measure of service concentration across all service lines, and specifies the extent to which the hospital focuses on particular specialties or is more balanced across specialties.

We also include controls for the inpatient activities excluded from our analysis relating to (i) unbundled activity and (ii) obstetric services. There are two options for this. First, we could control for the percentage of total volume that the excluded activity constitutes at a hospital trust t in a particular year t with $\mathbf{VolUnbund}_{th}$ and $\mathbf{VolObstetrics}_{th}$ for unbundled and obstetric services, respectively. Else, we could control for the percentage of total cost that the excluded activity constitutes with $\mathbf{CostUnbund}_{th}$ and $\mathbf{CostObstetrics}_{th}$. The results are consistent using either approach, with the results in this paper reported when using the volume controls.

We also note that some trusts operate multiple hospitals, meaning that activity may be distributed across multiple sites which can make measuring the scale and scope effects of interest challenging. To adjust for this, we include two further controls in the models. The first, $\mathbf{Sites}_{th}$, is a categorical variable equal to the number of acute and multi-service hospital sites that each trust operates. The second, $\mathbf{BedConc}_{th}$, is a control for the concentration of beds across the different hospital sites that each trust operates. This concentration measure is again based on the Herfindahl-Hirschman Index. In particular, if b_{tsh} is the proportion of total beds at hospital site s of trust h in year t, then the bed concentration at trust h is equal to $\sum_s b_{tsh}^2$. For example, if a hospital operates two 250-bed sites, then $b_{1h} = b_{2h} = 0.5$, and the concentration index is equal to $2 \times 0.5^2 = 0.5$.

Finally, we have included three other variables in the model: $\mathbf{Teach}_{th}$, which is a binary variable taking the value 1 if the hospital trust has teaching status and 0 otherwise, $\mathbf{Merger}_{th}$, which is a binary variable taking value 1 when the hospital trust was involved in a merger the previous year and 0 otherwise, and $\mathbf{Region}_h$, which indicates which of the 10 UK regions (so-called "strategic health authorities") the hospital belongs to.

To remain consistent with the MLM approach, all of the continuous controls (i.e. those that are not binary or categorical) are separated into their longitudinal (within-hospital) and cross-sectional (between-hospital) parts.

4. Discussion of Results

The within–between RE (MLM) regression models were estimated in R (version 3.3.3) using the `lmer()` function of the `lme4` package, with model parameters calculated using restricted maximum likelihood estimation (Bates et al. 2015). Recall that the unit of analysis for each regression model is a specialty in a hospital trust within a fixed admission category (elective or emergency), observed annually over a 10-year period.

Table 2 contains the most relevant regression output for costs and length of stay (LOS), separately for the two admission categories. The upper two panels report coefficient estimates and standard errors of the longitudinal and cross-sectional effects, respectively, of the four independent variables of interest. These coefficients capture direct economies of scale (the effect of increased volume in the focal specialty and focal admission category) and three spillover effects: (i) the effect of increased volume in *other specialties* in the focal admission category; (ii) the effect of increased volume in the *other admission category* in the focal specialty; (iii) the effect of increased volume in *other specialties* in the *other admission category.* The third panel ("Control structure") reports the factors that are included as fixed effects (FE) – indicated by a "Y" – and gives the estimated standard deviations (σ_z) of the factors that are modelled as random effects (RE). The lower panel ("Model fit") reports the marginal R^2, which describes the proportion of variance explained by non-random factors (e.g. the volume variables and controls) alone, and the conditional R^2, which describes the proportion of variance explained by both the non-random and random factors (Johnson 2014).

Before we discuss the results, we remind the reader that the cross-sectional effect coefficients refer to the effect of variation in time-averaged patient volumes between hospitals, while the longitudinal coefficients capture the effects of annual changes in patient volumes, above and beyond aggregate demand growth, which is controlled through year-fixed effects. The cross-sectional effects are therefore likely to capture cost-effects resulting from different asset structures afforded by hospitals with different average activity levels, the *hospital design effect*, while the longitudinal effects capture cost-effects of changes in asset utilization, in response to changing volume over time. Our focus is on the former effect, while controlling for the latter.

Since the dependent and independent variables have been log-transformed, their coefficients can be interpreted as elasticities, i.e., the coefficient is the percentage change in Cost (or LOS) associated with a 100% increase (i.e. doubling) of the respective annual volumes. Note that, as a consequence, the magnitude of a reported coefficient of the volume of the focal specialty (e.g. Elect. vol. (focal Sp)) is not directly comparable with the corresponding coefficient of the volume

Table 2 Model parameter estimates – MLMs using within-between volume decomposition

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.131^{***}	$0.007^{\dagger}$	-0.074^{***}	0.005
	(0.006)	(0.004)	(0.003)	(0.003)
Elect. vol. (other Sps)	-0.127^{***}	0.083^{**}	-0.028^{*}	0.104^{***}
	(0.028)	(0.026)	(0.014)	(0.024)
Emerg. vol. (focal Sp)	0.003	-0.177^{***}	0.012^{*}	-0.126^{***}
	(0.011)	(0.007)	(0.006)	(0.005)
Emerg. vol. (other Sps)	0.036	-0.181^{***}	$0.025^{\dagger}$	-0.110^{***}
	(0.027)	(0.025)	(0.014)	(0.024)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.048^{***}	0.031^{***}	-0.021^{***}	0.013^{**}
	(0.011)	(0.007)	(0.005)	(0.005)
Elect. vol. (other Sps)	0.048	0.137^{***}	0.013	0.054^{*}
	(0.039)	(0.030)	(0.016)	(0.026)
Emerg. vol. (focal Sp)	−0.012	-0.144^{***}	0.012	-0.106^{***}
	(0.019)	(0.011)	(0.008)	(0.008)
Emerg. vol. (other Sps)	−0.051	-0.110^{***}	0.007	−0.032
	(0.038)	(0.030)	(0.015)	(0.026)
Control structure				
Year	Y	Y	Y	Y
Specialty	Y	Y	Y	Y
Trust	0.080	0.072	0.030	0.065
Trust–year	0.084	0.091	0.042	0.093
Specialty–trust	0.147	0.088	0.058	0.066
Specialty–year	0.025	0.014	0.020	0.015
Residual std. error	0.209	0.140	0.105	0.092
Model fit				
Observations	20,057	21,507	20,057	21,507
Marginal R^2	0.127	0.215	0.144	0.152
Conditional R^2	0.519	0.626	0.458	0.724
Bayesian inf. crit.	1,758.5	-14,796.5	-26,284.0	-31,153.2

$^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; Inclusion of a FE in the control structure indicated by a "Y", inclusion of a RE indicated by the reporting of its estimated standard deviation.

of other specialties (e.g. Elect. vol. (other Sps)) because the total volume of all other specialties combined will be much larger than that of the single focal specialty. Hence a 100% increase of the former implies a much larger absolute increase than a 100% increase of the latter. To help the reader compare the coefficients, we calculate the effect of an increase by 1,000 patients per annum at the mean (as given in Table 1) and we report this in the final column (Column 4) of Table 3 which summarizes the estimated effects. We do so only for the more important cross-sectional effects of volume.

4.1. Economies of Scale and Spillover Effects for Elective Services

Starting from economies of scale within elective services (first column of Table 2), after controlling for asset utilization, we find that elective specialties structured to treat a higher volume of patients (the cross-sectional effect of volume) are associated with lower costs. More specifically a 10% increase in the average volume of elective patients within a specialty leads to a 0.48% ($p < 0.1\%$) reduction of cost per patient. In absolute terms, the marginal impact of increasing the average number of elective patients treated by a specialty by 1,000 per annum on the cost per patient is -3.7%. Turning to spillover effects from other elective specializations, after controlling for asset utilization, we find that the average volume of other elective specializations (cross-sectional volume effect) has no statistically significant effect on costs ($\beta = 0.048$, $p = 22.8\%$). Similarly, we find no volume-related spillover effect on elective costs from emergency patients either within the same medical specialization ($\beta = -0.012$, $p = 52.8\%$) or from different medical specializations and ($\beta = -0.051$, $p = 17.2\%$). These results are confirmed by the LOS regression in the third column of Table 2. In summary, the results of the cross-sectional differences in volume across hospital–specialties suggest that there exist economies of scale for elective care; services designed to treat a larger volume of elective patients generate costs savings. However, we find no evidence to suggest that the organizational integration of multiple specialty services, or the organizational integration of emergency and elective services, provide productivity benefits for elective services.

Turning to the longitudinal effects, which capture the impact of differential asset utilization within specialty, we find that elective costs are reduced as the annual volume of elective patients within the specialty increases ($\beta = -0.131$, $p < 0.1\%$) and as the annual volume of elective patients from other specialties increases ($\beta = -0.127$, $p < 0.1\%$)[6] but we find no statistically significant effects on costs from emergency patients volumes, either within ($\beta = 0.003$, $p = 78.3\%$) or between different specialties ($\beta = 0.036$, $p = 19.4\%$). The results on the impact of volume on LOS are similar both in direction and magnitude (with the only difference that some of the small coefficients that were not statistically significant at conventional levels for costs are marginally significant for LOS). The longitudinal results are consistent with the view that higher elective volume leads to higher utilization of assets designed for elective care which leads to a reduction of costs, but suggest that an increased volume of emergency patients confers no additional benefit. The latter is consistent with the observation that emergency patients have sufficiently differentiated medical needs from elective patients.

4.2. Economies of Scale and Spillover Effects for Emergency Services

Analogously to elective services, we find strong economies of scale in emergency services (second column of Table 2). After controlling for asset utilization, we find that a 10% increase in the average volume of emergency patients treated by a specialty reduces costs by 1.44% ($p < 0.1\%$). In addition to the positive economies of scale associated with an increase in the volume of emergency patients

[6] We remind the reader that even though the magnitude of these coefficients are comparable, the marginal effect of an additional patient within a specialty (the first effect) is much larger than the marginal effect of an additional patient from a different specialty (the second effect). As explained above, this is due to the fact these coefficients represent elasticities.

within a specialty, for emergency patients we also find a positive spillover effect associated with an increase in emergency volumes from other services. More specifically, after controlling for asset utilization, we find that a 10% increase in time-averaged emergency volume in other specialties to be associated with a 1.10% cost reduction. To help the reader compare the magnitude of these estimated effects, we note that the marginal impact of increasing the average number of emergency patients treated by the focal specialty (by other specialties) by 1,000 per annum on the cost per patient is -7.1% (-0.3%). The positive spillover from one medical specialty to another, present for emergency patients but not for electives, is consistent with the fact that emergency patients share more assets/resources (e.g. Emergency Department beds/physicians) across specialties than elective patients do.

In sharp contrast to elective services, the results suggest that there exists *negative* spillover effects from elective to emergency services. After controlling for asset utilization, we find that the cost of emergency patients increases when they are treated in hospitals designed to cater for a larger volume of elective patients. More specifically, after controlling for asset utilization, a 10% increase in the elective patient volume of the focal specialty (other specialties) is associated with an increase in emergency costs by 0.31%, $p < 0.1\%$ (1.37%, $p < 0.1\%$) in the focal specialty. The associated marginal effect of increasing the average number of elective patient volume of the focal specialty (other specialties) by 1,000 per annum on the emergency costs of the focal specialty is 2.4% (0.4%).

Together with the results of the previous section, these findings suggest that the spillover effect of volume of emergencies and electives on each other is asymmetric. If elective volume increases, the optimization of hospital assets and patient pathways shifts away from emergencies towards electives leading to a reduction of costs for the former at the expense of productivity loss at the latter. The opposite is not true – as emergency volume increases, elective pathways and assets used for treating elective patients appear to be protected – a result consistent with calls to ring-fence elective care in order to improve productivity (Kjekshus and Hagen 2005).

Turning to the longitudinal effects, which capture the impact of differential asset utilization within a specialty, we find that the cost of treating emergency patients is reduced as the annual volume of emergency patients within the specialty increases ($\beta = -0.177$, $p < 0.1\%$) and as the annual volume of emergency patients from other specialties increases ($\beta = -0.181$, $p < 0.1\%$). In addition, we find some evidence that an increase in annual volume of elective patients either within ($\beta = 0.007$, $p = 9.57\%$) or across specialties ($\beta = 0.083$, $p = 0.13\%$). Together, the longitudinal effects are consistent with the more important cross-sectional effects.

As in the case of elective services, the results from the LOS regressions are similar in both direction and magnitude and confirm both the positive economies of scale as well as the negative spillovers from elective to emergency services (see the fourth column of Table 2).

Table 3 Marginal effects at the mean

Effect on...	of an increase in...	from the...	Approximate marginal effect size on costs[(1)]
Elective productivity	Elective vol.	Focal Sp	−3.7%
		Other Sps	–
	Emergency vol.	Focal Sp	–
		Other Sps	–
Emergency productivity	Elective vol.	Focal Sp	+2.4%
		Other Sps	+0.4%
	Emergency vol.	Focal Sp	−7.1%
		Other Sps	−0.3%

[(1)]Effect on costs is approximated by adding 1,000 patients per annum (from the specialty(s) and admission category in the corresponding row) to the mean volume level given in Table 1. The effect is based on the cross-sectional-volume effects estimated in Table 2.

5. Limitations, Robustness Tests, and Alternative Specifications

As with all "multi-firm" studies based on accounting costs, our analysis has limitations due to the unobserved degree of adherence of individual hospital cost accounting systems to the national guidelines. We believe that the aggregation of the granular HRG codes to which costs are allocated to the coarser level of HRG chapters as the unit of analysis helps alleviate this problem as accounting inaccuracies within specialties average out at the aggregate level and accounting misallocations between specialties are less likely. In addition, we corroborate our findings with an analysis of LOS, which is unaffected by hospital accounting systems but highly correlated with costs, and which confirms our results.

Nevertheless, to investigate the robustness of the results presented in the previous section, we extend the empirical model to allow the volume effects to vary by specialty, discuss potential reverse causality, and describe the findings from a number of other model specifications. More details on these additional analyses are presented in the supplementary material. Throughout these sections the emphasis of the discussion is on the more important cross-sectional volume effects, but we also note that the results of the longitudinal volume remain similar.

5.1. Heterogeneous Effects Across Specialties

In the models presented in the previous section we estimated the average impact of volume on costs and LOS across different specialties, implicitly assuming that this impact of volume was homogeneous across the different specialties. We can relax this assumption and allow for heterogeneous slopes for each of the specialties. To do this we estimate a model where, in addition to random intercepts, we *also* allow for random slopes. In essence these random slopes allow specialty specific deviations from the common "overall" volume effect. This is a more flexible approach than adding interaction terms between the specialties and volume effects of interest, although the interpretation is similar. We discuss and present results here for random slope estimates for Elect. vol. (focal Sp) and Emerg. vol. (focal Sp), with results for volume in other specialties being similar and reported in Section EC.1 in the supplementary material. Exact details of how the random slopes are implemented can also be found in the supplementary material.

Figure 2 shows the random slope estimates of the between-effects in the cost models, together with bootstrapped 95% confidence intervals (using 10,000 simulations from the posterior distribution). The separate slopes that are derived for each specialty give an estimate of the specialty-specific between-effects of elective and emergency volume on cost. These can be compared with the combined slope estimates from Section 4, which are also plotted (as "ALL") in Figure 2. Comparing these, it can be seen that the directions of the specialty-specific effects are consistent with the combined estimates, with 95% confidence intervals overlapping in nearly all cases.[7] Due to limited data, the confidence intervals are wide for these specialty-dependent random slopes, and so in presenting the main results we prefer to report the aggregate effects across specialties.

One limitation of the work presented above is that there may be certain specialties that share more resources than others. Since our empirical strategy combines all other elective specialties, our results may underestimate the potential economies of scope that might be achieved through combining particular elective specialties. We note that this does not invalidate our findings of economies of scale within a specialty, nor the negative spillover effect from electives to emergencies. Instead, it suggests that there may be even further cost savings that might be achieved through being more strategic, e.g., by growing electives of related specialties.

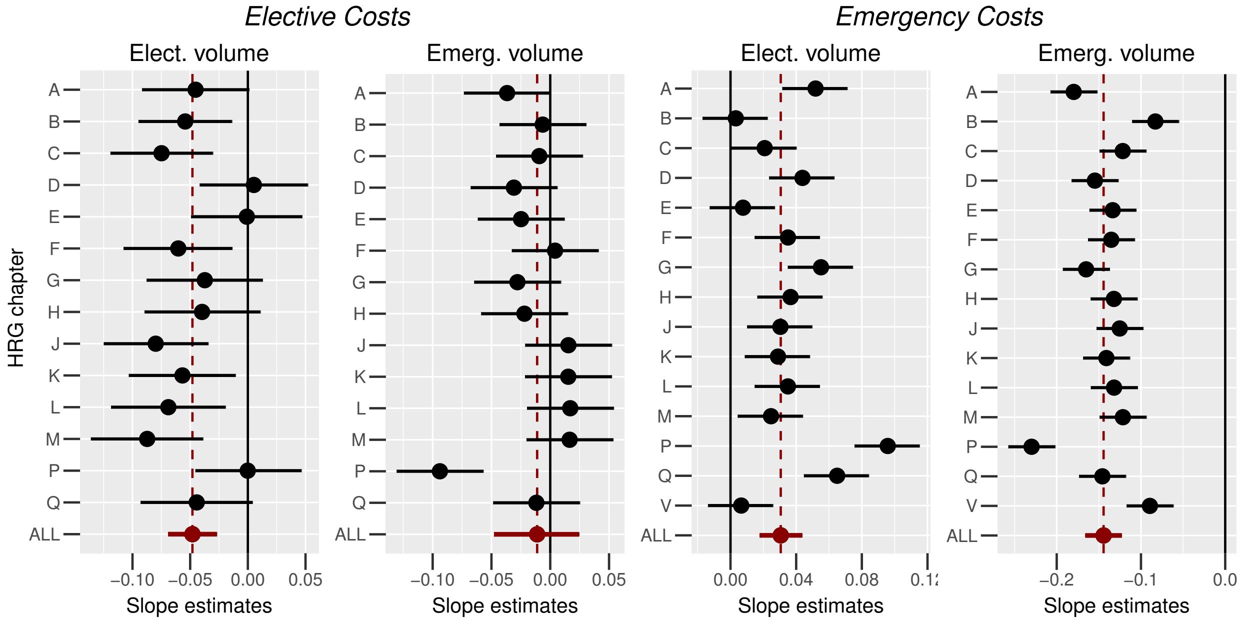
5.2. Reverse Causality

In this paper we have argued that higher volumes confer a productivity advantage. However, the direction of causality is not apparent: It could be argued instead that the positive relationship identified between volume and productivity is actually the result of either (i) more cost-effective hospitals being referred or taking action to attract a higher volume of patients or (ii) patients self-selecting these hospitals. Here we discuss both of these alternatives and combine empirical arguments made using the data with other evidence to suggest that this is not the case.

First, we consider whether patients are referred more often to more productive hospitals or if those hospitals use their stronger financial position to take action (e.g. through marketing or

[7] Observe that specialty P, corresponding to pediatrics, appears not to follow the general trend. This may not be too surprising since pediatrics is a highly specialized service for which only a limited number of hospitals provide treatment across the full spectrum of possible conditions. We note that excluding this specialty from the analysis does not change the qualitative findings.

Figure 2 **Random slope coefficient estimates for the effect of volume in the focal specialty on costs, reported by specialty (black) and combined (red), with bootstrapped 95% confidence intervals.**



Note. A - nervous system; B - eyes & periorbita; C - mouth, head, neck, & ears; D - respiratory system; E - cardiac surgery & primary cardiac conditions; F - digestive system; G - hepatobiliary & pancreatic system; H - musculoskeletal system; J - skin, breast & burns; K - endocrine & metabolic system; L - urinary tract & male reproductive system; M - female reproductive system; P - diseases of childhood and neonates; Q - vascular system; V - multiple trauma.

lobbying) to increase their patient pool. We note that any effect is likely to be small, since a recent study by the King's Fund, an independent UK-based healthcare think tank, found that most hospital trusts operated in a defined geographical market and only competed for patients "at the boundaries of their catchment areas, where another provider was equidistant" (Dixon et al. 2010). Nevertheless, to test this we run additional analysis where we examine whether past financial performance is a predictor of future patient volumes. If better performing hospitals are able to attract or are referred a higher volume of patients, then we would expect lower costs in the past to be positively correlated with higher patient volume in the future. However, our regression results (reported in Section EC.4.1 of the supplementary material) suggest that, if anything, the opposite occurs: hospitals that are higher cost in the past are *more* rather than less likely to increase patient volumes than lower cost hospitals.

Next, we consider whether patients self-select more productive hospitals. We note first that as health services in the UK are free at the point of care there is little incentive for a patient to select their care provider based on cost. Indeed, such information is not made readily available. However, while patients are unlikely to decide based on cost, it is possible that they will select based on quality. As cost and quality are often correlated, and quality is an unobserved factor that we do not account for in this analysis, this could be driving the results. Information on the quality of hospitals, however, has not been readily available until recently, and it remains challenging for patients to compare treatment for procedures at different hospitals. Patients may infer quality through other more tacit means, however, e.g. by way of word of mouth. To test this, we utilized data from a government administered Adult Inpatient Satisfaction Survey (NHS 2017). This annual survey contains responses to various questions about patients' experiences at every acute NHS trust, and is available over the same 10 year period as the cost data. The responses are aggregated into an Overall Patient Experience Score which serves as an excellent proxy for perceived quality and so we would expect to capture much of the word of mouth effect. When introduced into the MLMs this variable has little to no impact and our main results remain unchanged (see Section EC.4.2 of the supplementary material). This is consistent with past research that has shown that there is little, if any, evidence of patients (or their physicians) exercising such choice (e.g. Gaynor et al. 2004, Gowrisankaran et al. 2006).

We also address the reverse causality concerns by re-running the analysis using a subset of the data corresponding to those hospital trusts that are geographically more isolated, with a restriction that the nearest trust can be no closer than 20km away. This has the effect of removing all hospital trusts located in cities and other more densely populated regions and, thus, reducing the number of trust-year observations by 64%, from 1,434 to 517. While this does not entirely avoid the problem of selection, the selection effect should be weaker in this subsample (as it is more inconvenient for a patient to attend another provider and hospitals have less ability to increase patient intake), especially for emergency patients, who need to be treated quickly. Therefore, if reverse causality were driving our results, then we would expect to find weaker evidence of productivity improvements from pooling similar types of activity when using this sample. The results (available in Section EC.4.3 of the supplementary material) show that this is not the case, with coefficient estimates nearly identical in sign and scale.

Together, this evidence suggests that the effects identified are very unlikely to be the result of reverse causality.

5.3. Other Robustness Checks and Modeling Alternatives

Another plausible type of endogeneity is selection by hospitals: Certain hospitals may choose to offer a subset of elective and/or emergency services (i.e. treat patients with a subset of conditions/HRGs only), and the choice of which services they offer may well depend on the profitability of these services. We have already partially accounted for this in our models by controlling for hospital–specialty effects as well as for the proportion of services, $\mathbf{Prop}_{thCp}$, offered by a hospital within each specialty in each year. Nevertheless, if specialties were formed endogenously in the way described above, then we might expect hospitals that offer fewer services also to be more profitable. In Section EC.5 of the supplementary material we show that there is little evidence of endogenous selection for emergency patients. For elective patients, we find that those hospitals that operate at higher volumes are less, not more, selective and offer a greater variety of services. If endogenous specialty formation were driving our results, we would, therefore, expect to find effects in the opposite direction to those we observe.

One concern when working with panel data is that errors may be autocorrelated, leading to underestimation of the standard errors of the estimated coefficients when autocorrelation is positive and potentially biasing the estimated coefficients in the within–between formulation (Hsiao 2015). We perform formal hypothesis testing with the Baltagi–Wu LBI test statistic and also extend our MLMs to allow the error term to be first-order autoregressive, i.e. to have AR(1) disturbances. Although, unsurprisingly, there exists some evidence of autocorrelation, the results remain consistent in terms of sign, scale and significance when we adjust our models to account for this effect (refer to Section EC.2 of the supplementary material for further details).

One might also be concerned about the high correlation between the various cross-sectional volume measures. To explore this further, we re-ran analysis but dropped each volume measure from the model one-at-a-time. Note that this approach has limitations since we trade-off multicollinearity concerns with a potential omitted variable bias that may arise from dropping a significant explanatory variable. These models show all of the findings to hold, except for the effect of emergency volume from other specialties on emergency costs in the focal specialty. Further testing for evidence for multicollinearity suggests this is not a major concern, i.e. all generalized VIFs take value less than 5.

Another possibility we consider in Section EC.3 of the supplementary material is that there may be non-linear effects of volume on costs. Although the models we estimate are already non-linear (as they involve the logarithmic transformations of both the dependent and independent variables) and suggest diminishing returns to scale (as the estimated coefficients are all < 1 and > -1), we also estimate models in which we add a squared-volume term for each of the cross-sectional effects. We find no evidence of any additional non-linear effect (reported in Section EC.3 of the supplementary material).

In addition to the models discussed above, we estimate a number of alternative model specifications that (i) cap costs at the HRG level to reduce the influence of outliers, capping below at 1/5th and above at 5 times the system-wide median, (ii) only compare costs for a subset of HRGs for which treatment in each year is provided in at least 80% of the hospital trusts in the sample, and (iii) constrain the sample to only include those specialty–trusts with a minimum volume level (e.g. >25% of the system-wide median) in order to reduce the potential influence of outliers. Since some trusts operate multiple hospital sites (with typically one large, main hospital and one or more smaller hospital sites), we also repeat the analysis for the subset of trusts with a single hospital site. Finally, we examine whether there is evidence of asset changes over time in hospitals by re-running the analysis allowing for one major structural change during the sample period per hospital trust. The results of these estimations are reported in Section EC.6 of the supplementary material and are qualitatively and quantitatively similar to those in Section 4 of this paper.

6. Managerial and Policy Implications

From a productivity perspective, the prevailing model of the fully comprehensive general hospital is predicated on the assumption that there are economies of scale and scope that come from pooling planned (elective) and unplanned (emergency) patient services and from pooling different specialties. However, our data and analysis suggest that there are in fact significant diseconomies associated with the pooling of elective and emergency patients. Specifically, an increase in elective services is associated with a significant productivity drop in emergency services. Furthermore, while the collocation of different specialties provides economies of scope for emergency patients, there is no evidence of positive productivity spillovers between specialties for elective patients. These findings have important implications for hospital growth strategies and the configuration of hospital systems, which we explore further in this section through counterfactual analyses.

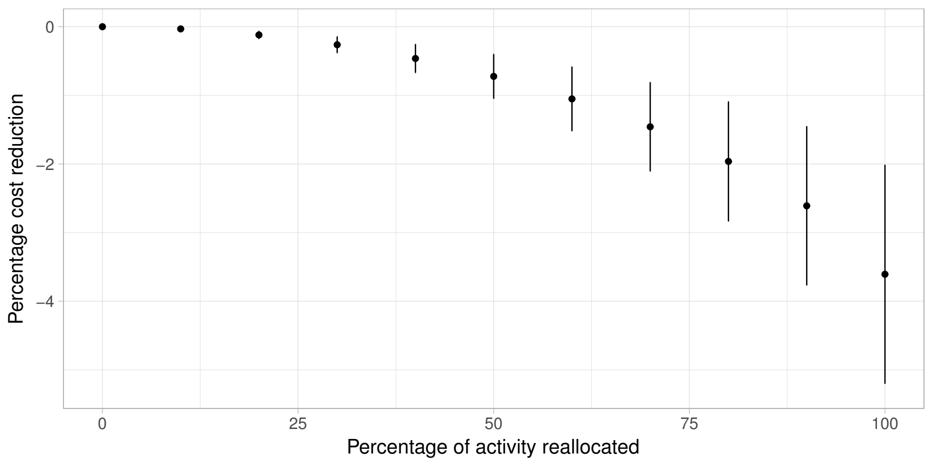
First, when hospitals consider growth strategies they have to be aware that while increasing elective activity improves the productivity of their elective patients, it has a negative impact on emergency activity, not only within the specialty that is growing but also for emergency patients in other specialties. To illustrate this, consider the model-predicted effect of different growth strategies for a major London hospital, St. George's, which admitted about 117,500 elective and emergency patients in 2015/16 in our dataset at a total cost of $\sim$ £220m. We estimate the impact of increasing total patient admissions to 141,000 per annum as a result of one of three strategies: (i) a 20% expansion across the board in elective and emergency volume, (ii) a 33% increase in emergency activity only, or (iii) a 50% increase in elective volume only, where growth causes the volume in each specialty to increase by the same percentage. Using the modeling results from Section 4, and focusing on the cross-sectional effect associated with different asset configuration rather than higher utilization of existing assets, we estimate that in the first scenario, elective costs would fall by 0.9% and emergency costs by 1.6%, leading to a total cost saving of £3m per annum. The emergency growth strategy would not affect elective costs but would reduce emergency costs by

7.3%, leading to a total cost saving of £11m per annum. Finally, the elective growth strategy would reduce elective costs by 2.0% but would have the unintended consequence of a 6.7% *increase* in emergency costs, leading to a total cost *increase* of £7m per annum. The negative spillover across all emergency services quickly erodes the productivity benefits of higher volume in elective services. This finding is surprising and important: The majority of hospitals in the UK are in deficit in the 2015/16 financial year and most chief executives see growth in elective activity, which is easier to plan and has less variation in costs, as the preferred way of increasing productivity to turn their hospital around. Few hospital managers would consider expanding their emergency activity. From a cost-management perspective, our results suggest that an elective growth strategy can be counterproductive if the hospital has high emergency volume and that in order to reduce costs it may actually be better to increase emergency activity instead.

The second important implication of our findings relates to regional hospital systems. Our results suggest that removing elective volume from general hospitals and instead treating these patients in regional *focused factories* should improve productivity for both the re-routed elective patients and the emergency patients remaining in the downsized general hospitals. To investigate the possible cost savings at the regional level, we present the results of a counterfactual analysis based on a plausible re-organization of elective services in London. We assume that any two hospital trusts in the city might agree to redistribute their elective services in such a way that there is no duplication of specialties between the two hospitals. We then estimate the cost implications arising from the increase in elective volume within each specialty. To minimize the need for additional capacity investment, we match hospital trusts pairwise based on their size, with the match made by pairing trusts that are most similar in terms of their total elective volume. Using the new allocation and the cross-sectional volume effects reported in §5, we calculate that for the trust-years in our analysis the total cost of providing elective care would be reduced by 3.6% (from £11.22bn to £10.81bn) per annum. If instead we only move 10%, 20%, 30% etc. of the activity then lesser gains can be achieved, as shown in Figure 3. Note that the cost savings could potentially be greater if (i) more than two hospital trusts worked together and (ii) the reallocation was based not only on volume but also on costs (so that the increased elective volume would be routed to the cheapest hospital). This finding implies that even simple regional reorganization may result in substantial cost savings.

Our findings also reconcile two seemingly opposing trends: (1) for small general hospitals to be closed or downgraded to urgent care centers and activity moved to larger general hospitals in the proximity and (2) for greater specialization with the opening of specialist hospitals focusing on only particular types of conditions. Interestingly, we show that these trends may not be at odds

Figure 3 Percentage reduction in total cost (with 95% confidence intervals) of elective activity in London when a percentage of elective activity is reallocated between two trusts.



and that the cost of providing care to different types of patients may be reduced through these different approaches. In particular, the productivity of elective care would benefit if elective patients were treated in specialist hospitals or regional treatment centers focused on specific specialties. We estimate, for example, that if London were to operate 14 such *focused factories* for each of the 14 specialties, then costs could be reduced further to £9.76bn: a saving of 13.6%. In addition, emergency patients would benefit from being treated in large, general acute hospitals that focus primarily on emergency care and treat a full spectrum of services. Implementing different service delivery modes for planned and unplanned activity could, therefore, be a highly effective way of increasing the productivity (and quality – see e.g. RCS/DH 2007, Kuntz et al. 2017) of hospital services in the longer term. Despite the large productivity gains suggested by these counterfactual analyses, there may well be reasons beyond the scope of this study – such as quality of care, patient and physician preferences, hospital teaching needs – for which such dramatic redesigns may not be practical to implement. We also acknowledge that this work has not been able to uncover the exact mechanisms that give rise to the positive and negative spillover effects identified in the paper. Future research, using more detailed data than currently available, should look to address this.

References

AHRQ (2014) National healthcare disparities report. Technical Report AHRQ Publication No. 14-0006, Agency for Healthcare Research and Quality.

Aletras V (1997) Part II: The relationship between volume and the scope of activity and hospital costs. *NHS Centre for Reviews and Dissemination* 8(2).

Argote L (1982) Input uncertainty and organizational coordination in hospital emergency units. *Administrative Science Quarterly* 27(3):420–434.

Argote L (2013) *Organizational learning: Creating, retaining and transferring knowledge* (New York: Springer Science & Business Media), 2nd edition.

ASGBI (2007) Emergency general surgery: The future. Technical report, Association of Surgeons of Great Britain and Ireland, London.

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48.

Batt R, Terwiesch C (2017) Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science* 63(11):3531–3551.

Bell A, Jones K (2015) Explaining fixed effects: Random effects modeling of time-series cross- sectional and panel data. *Political Science Research and Methods* 3(1):133–153.

Boh W, Slaughter S, Espinosa J (2007) Learning from experience in software development: A multilevel analysis. *Management Science* 53(8):1315–1331.

Bohmer R (2009) *Designing care: Aligning the nature and management of health care* (Boston, MA: Harvard Business School Press).

Carey K, Burgess JF, Young GJ (2015) Economies of scale and scope: The case of specialty hospitals. *Contemporary Economic Policy* 33(1):104–117.

Certo ST, Withers M, Semadeni M (2017) A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal* 38(7):1536–1556.

Chan C, Farias V, Escobar G (2017) The impact of delays on service times in the intensive care unit. *Management Science* 63(7):2049–2072.

Christensen C, Grossman J, Hwang J (2009) *The Innovator's Prescription: A Disruptive Solution for Health Care* (New York: McGraw-Hill).

Clark J (2012) Comorbidity and the limitations of volume and focus as organizing principles. *Medical Care Research and Review* 69(1):83–102.

Clark J, Huckman R (2012) Broadening focus: Spillovers, complementarities, and specialization in the hospital industry. *Management Science* 58(4):708–722.

DH (2013) A simple guide to payment by results. Technical report, Department of Health, UK, URL `https://www.gov.uk/government/publications/simple-guide-to-payment-by-results`, Accessed: 2016-03-09.

DH (2016) Reference costs 2015-16. Technical report, Department of Health, UK, URL `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/577083/Reference_Costs_2015-16.pdf`, Accessed: 2017-05-15.

Dijk N, Sluis E (2004) To pool or not to pool in call centers. *Production and Operations Management* 17(3):296–305.

Dixon A, Appleby J, Robertson R, Burge P, Devlin N, Magee N (2010) *Patient choice: How patients choose and how providers respond* (London, UK: The King's Fund).

Dranove D (1998) Economies of scale in non-revenue producing cost centers: Implications for hospital mergers. *Journal of Health Economics* 17(1):69–83.

Fetter R (1991) Diagnosis related groups: Understanding hospital performance. *Interfaces* 21(1):6–26.

Fisher M, Ittner C (1999) The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Science* 45(6):771–786.

Freeman M, Robinson S, Scholtes S (2017a) Gatekeeping under congestion: An empirical study of referral errors in the emergency department, Working paper, 2017/59/TOM, INSEAD.

Freeman M, Savva N, Scholtes S (2017b) Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science* 63(10):3147–3167.

Gaughan J, Mason A, Street A, Ward P (2012) English hospitals can improve their use of resources: An analysis of costs and length of stay for ten treatments. Technical report, Centre for Health Economics, University of York, London, CHE Research Paper 78.

Gaynor M, Seider H, Vogt W (2004) Volume-outcome and antitrust in U.S. health care markets, Unpublished manuscript, Carnegie Mellon University.

Gaynor MS, Kleiner SA, Vogt WB (2015) Analysis of hospital production: An output index approach. *Journal of Applied Econometrics* 30(3):398–421.

Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models* (New York, NY, USA: Cambridge University Press).

Giancotti M, Guglielmo A, Mauro M (2017) Efficiency and optimal size of hospitals: Results of a systematic search. *PLoS One* 12(3).

Gimeno J, Woo C (1999) Multimarket contact, economies of scope, and firm performance. *Academy of Management Journal* 42(3):239–259.

Gowrisankaran G, Ho V, Town R (2006) Causality, learning and forgetting in surgery, Unpublished manuscript.

Green L, Savin S, Savva N (2013) "Nursevendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Science* 59(10):2237–2256.

Greenwald L, Cromwell J, Adamache W, Bernard S, Drozd E, Root E, Devers K (2006) Specialty versus community hospitals: Referrals, quality, and community benefits. *Health affairs* 25(1):106–118.

Henderson R, Cockburn I (1996) Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *The RAND Journal of Economics* 27(1):32–59.

HFMA (2016) Acute health clinical costing standards. Technical report, Healthcare Financial Management Association, URL `https://www.hfma.org.uk/docs/default-source/our-work/costing/Clinical-Costing-Standards/acute-standards-201617.pdf`, Accessed: 2016-03-09.

Hill CW, Hoskisson RE (1987) Strategy and structure in the multiproduct firm. *Academy of Management Review* 12(2):331–341.

HMT (2015) HMT public expenditure statistical analyses (PESA). Technical report, HM Treasury, UK, URL `https://www.gov.uk/government/collections/public-expenditure-statistical-analyses-pesa`, Accessed: 2016-03-12.

Hopp W, Lovejoy W (2012) *Hospital operations: Principles of high efficiency health care* (FT Press).

Hopp W, Spearman M (2004) To pull or not to pull: What is the question? *Manufacturing & Service Operations Management* 6(2):133–148.

HSCIC (2015) HRG4 2014/15 consultation grouper roots, Health & Social Care Information System, Accessed: 2016-03-09.

Hsiao C (2015) *Analysis of panel data* (Cambridge, UK: Cambridge University Press), 3rd edition.

Huckman R, Pisano G (2006) The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science* 52(4):473–488.

Johnson P (2014) Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution* 5(9):944–946.

Joustra P, Van der Sluis E, Van Dijk N (2010) To pool or not to pool in hospitals: A theoretical and practical comparison for a radiotherapy outpatient department. *Annals of Operations Research* 178(1):77–89.

Junius K (1997) Economies of scale: A survey of the empirical literature, Working paper, No. 813, Institut für Weltwirtschaft (IfW), Kiel.

KC D, Staats B (2012) Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing & Service Operations Management* 14(4):618–633.

KC D, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

KC D, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* 14(1):50–65.

Kekre S, Srinivasan K (1990) Broader product line: a necessity to achieve success? *Management Science* 36(10):1216–1232.

Kim H (1987) Economies of scale in multi-product firms: an empirical analysis. *Economica* 54(214):185–206.

Kim S, Chan C, Olivares M, Escobar G (2014) ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* 61(1):19–38.

Kjekshus L, Hagen T (2005) Ring fencing of elective surgery: does it affect hospital efficiency? *Health Services Management Research* 18(3):186–197.

Kristensen T, Olsen K, Kilsmark J, Pedersen K (2008) Economies of scale and optimal size of hospitals: Empirical results for Danish public hospitals, Unpublished manuscript, University of Southern Denmark.

Kuntz L, Mennicken R, Scholtes S (2014) Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* 61(4):754–771.

Kuntz L, Scholtes S, Sülz S (2017) Separate and concentrate: Accounting for patient complexity in general hospitals. *Management Science (Forthcoming)* .

Marini G, Miraldo M (2009) Economies of scale and scope in the English hospital sector, Unpublished manuscript, University of York.

McDermott CM, Stock GN, Shah R (2011) Relating focus to quality and cost in a healthcare setting. *Operations Management Research* 4(3–4):127–137.

Monitor (2012) Economies of scale and scope in healthcare markets. Technical report, URL https://www.gov.uk/government/consultations/healthcare-markets-economies-of-scale-and-scope, Accessed: 2017-07-14.

Monitor (2013) A guide to the market forces factor. Technical report, Publication code: IRG 31/13.

Monitor (2015) Improving productivity in elective care. Technical report, URL https://www.gov.uk/guidance/improving-productivity-in-elective-care, Accessed: 2016-03-09.

Moore F (1959) Economies of scale: Some statistical evidence. *The Quarterly Journal of Economics* 73(2):232–245.

Mundlak Y (1978) On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society* 46(1):69–85.

NAO (2011) Delivering efficiency savings in the NHS. Technical report, National Audit Office, UK.

Narayanan S, Balasubramanian S, Swaminathan J (2009) A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Science* 55(11):1861–1876.

Nasr A, Reichardt K, Fitzgerald K, Arumugusamy M, Keeling P, Walsh T (2004) Impact of emergency admissions on elective surgical workload. *Irish Journal of Medical Science* 173(3):133.

NHS (2017) Overall patient experience scores. Technical report, Analytical Services, NHS England, URL https://www.england.nhs.uk/statistics/statistical-work-areas/pat-exp/, Accessed: 2017-07-17.

NT (2016) NHS in numbers. Technical report, Nuffield Trust, URL http://www.nuffieldtrust.org.uk/nhs-numbers-0, Accessed: 2016-03-12.

Panzar J, Willig R (1977) Economies of scale in multi-output production. *The Quarterly Journal of Economics* 91(3):481–493.

Panzar J, Willig R (1981) Economies of scope. *American Economic Review* 71(2):268–272.

Penrose E (1959) *The Theory of the Growth of the Firm* (New York: John Wiley).

Pisano G, Bohmer R, Edmondson A (2001) Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Science* 47(6):752–768.

Porter M (1985) *The Competitive Advantage: Creating and Sustaining Superior Performance* (Free Press).

Posnett J (2002) Are bigger hospitals better? McKee M, Healy J, eds., *Hospitals in a changing Europe*, chapter 6, 100–118 (Buckingham: Open University Press).

Powell A, Savin S, Savva N (2012) Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing & Service Operations Management* 14(4):512–528.

Preyra C, Pink G (2006) Scale and scope efficiencies through hospital consolidations. *Journal of Health Economics* 25(6):1049–1068.

Ramdas K, Saleh K, Stern S, Liu H (2017) Variety and experience: Learning and forgetting in the use of surgical devices. *Management Science* 64(6):2590–2608.

Rawley E, Simcoe T (2010) Diversification, diseconomies of scope, and vertical contracting: Evidence from the taxicab industry. *Management Science* 56(9):1534–1550.

RCS/DH (2007) Separating emergency and elective surgical care: Recommendations for practice. Technical report, The Royal College of Surgeons of England, Department of Health, London.

RCS/DH (2010) The higher risk surgical patient: Towards improved care for a forgotten group. Technical report, The Royal College of Surgeons of England, Department of Health, London.

Risse G (1999) *Mending bodies, saving souls: A history of hospitals* (Oxford University Press).

Sav G (2004) Higher education costs and scale and scope economies. *Applied Economics* 36(6):607–614.

Savva N, Tezcan T, Yildiz O (2018) Can yardstick competition reduce waiting times? *Management Science (Forthcoming)* .

Schilling M, Vidal P, Ployhart R, Marangoni A (2003) Learning by doing something else: Variation, relatedness, and the learning curve. *Management Science* 49(1):39–56.

Schoar A (2002) Effects of corporate diversification on productivity. *The Journal of Finance* 57(6):2379–2403.

Shleifer A (1985) A theory of yardstick competition. *The RAND Journal of Economics* 319–327.

Silk A, Berndt E (1993) Scale and scope effects on advertising agency costs. *Marketing Science* 12(1):53–72.

Skinner W (1974) The focused factory. *Harvard Business Review* 113–121.

Staats B, Gino F (2012) Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Science* 58(6):1141–1159.

Tsikriktsis N (2007) The effect of operational performance and focus on profitability: A longitudinal study of the us airline industry. *Manufacturing & Service Operations Management* 9(4):506–517.

UKAC (2011) Improving coding, costing and commissioning: Annual report on the payment by results data assurance programme 2010-11. Technical report, Audit Commission UK, URL `http://collections.europarchive.org/tna/20121205001956/http://audit-commission.gov.uk/SiteCollectionDocuments/Downloads/pbrannualreport2011.pdf`, Accessed: 2016-03-09.

Vanberkel P, Boucherie R, Hans E, Hurink J, Litvak N (2012) Efficiency evaluation for pooling resources in health care. *OR Spectrum* 34(2):371–390.

e-companion to "Economies of Scale and Scope in Hospitals"

This e-companion contains supporting material designed to accompany the investigation presented in the main paper. In Section EC.1 we provide random slope estimates for the effects of volume from *other* specialty, to augment those provided for volume of the *same* specialty given in the Section 5.1 of the paper. In Section EC.2 we show that there is no evidence that the errors are autocorrelated. In Section EC.3 we investigate the possibility of non-linear volume effects, and find little evidence to suggest this is the case. In Section EC.4 we discuss and argue against the possibility that our findings are due to reverse causality. In Section EC.5 we discuss the fact that elective specialties might be formed endogeneously based on financial viability, and show how we account for this, provide additional robustness checks, and discuss how – if anything – this would be expected to work against our findings. In Section EC.6 we report on the results a number of additional tests that (i) are performed on a subset of data corresponding to hospital trusts that are more geographically isolated, (ii) limit the possibility of extreme cost outliers driving the results, (iii) compare hospital trusts based on a set of common (rather than all) HRGs that are performed by most (>80%) of trusts, (iv) re-run the models on a subset of the specialties for which hospital trusts treat a high enough volume of patients, and (v) restrict the sample to trusts that operate only a single hospital site. The results from all of the models in Section EC.6 are in line with those reported in the paper. In Section EC.7 we combine the elective and emergency panels are report results from a joint analysis which allows for the errors terms across the two patient types to be correlated. In Section EC.8 we present more details on how we generated the dependent variables used in the main analysis. In Section EC.9 we provide a discussion on the longitudinal and cross-section effect of volume. Finally, in Section EC.10, we provide an in-depth literature review.

EC.1. Random slopes – hospital trust volume effects

In Section 5.1 of the paper we report on random slopes estimates for the effect of same–speciality volume on hospital trust costs. First we must discuss how these effects were estimated, before extending them to examine whether the spillover effect of volume from the *other* specialties on cost of the focal specialty differs by specialty.

To estimate the random slopes in Section 5.1 of the paper we include in Equation (3) random specialty-dependent slopes $\beta^{LT}_{1,(C)[i]}$, $\beta^{LT}_{2,(C)[i]}$, $\beta^{CS}_{1,(C)[i]}$ and $\beta^{CS}_{2,(C)[i]}$, respectively. These specialty-dependent random slopes model the degree to which the volume effect for a given specialty deviates from the *global* volume effect.

It is typical in the MLM literature to allow the random slopes to be correlated with the specialty-specific intercepts. To achieve this we need to also replace the specialty FE, $P_{(C)[i]}$, in Equation (4) – which we use in place of a RE (see footnote 5) – with a RE, $\alpha_{(C)[i]}$. We then model $(\alpha_{(C)[i]}, \beta^{LT}_{1,(C)[i]}, \beta^{LT}_{2,(C)[i]}, \beta^{CS}_{1,(C)[i]}, \beta^{CS}_{2,(C)[i]})$ using a multivariate normal distribution to allow for correlation between the REs (see Gelman and Hill 2007, for details). This requires the estimation of 15 parameters: five variance terms, one for each of the random slopes, plus ten pairwise correlation terms between each of the random slopes. While these models result in slightly improved model fit (the BIC is reduced from $1,758.5$ to $1,588.4$ for the elective cost MLM, and from $-14,796.5$ to $-14,972.1$ for the emergency cost MLM), we note that the global effects remain almost identical in terms of sign, size, and significance.

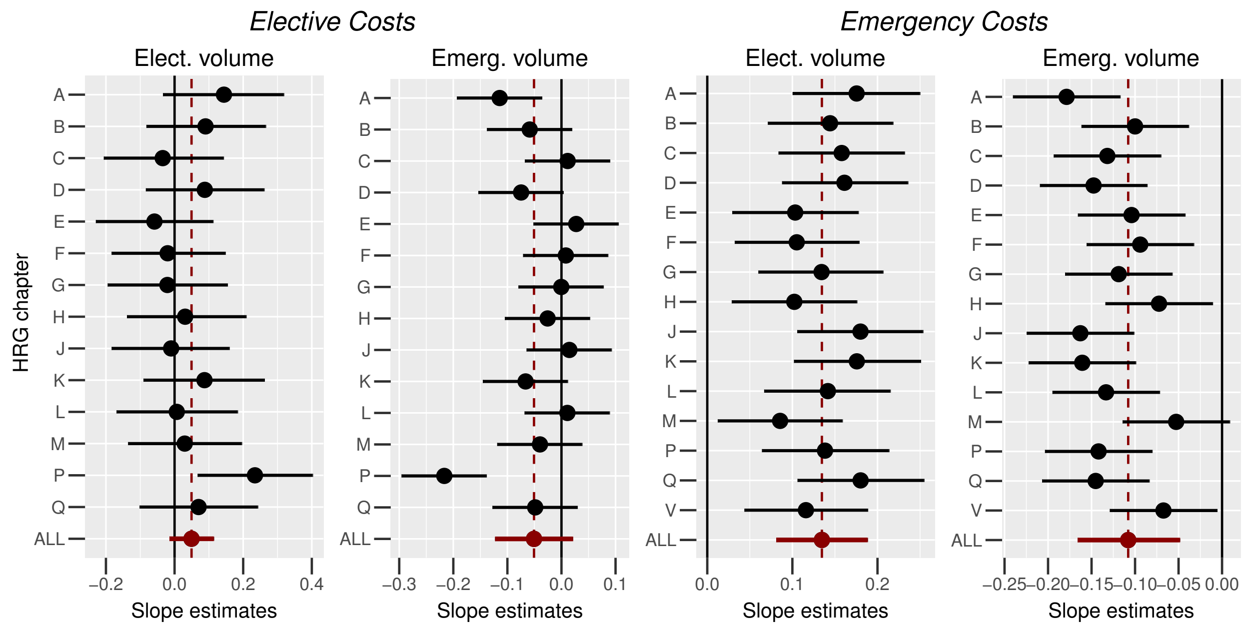
In order to identify the spillovers effects of volume from *other* specialties onto the focal specialty we can re-run the above analysis but where we instead include in Equation (3) random specialty-dependent slopes $\beta^{LT}_{3,(C)[i]}$, $\beta^{LT}_{4,(C)[i]}$, $\beta^{CS}_{3,(C)[i]}$ and $\beta^{CS}_{4,(C)[i]}$, respectively. We model the random slopes jointly as a multivariate normal distribution, as above. The results are plotted – together with bootstrapped 95% confidence intervals using 10,000 simulations from the posterior distribution of the MLMs – in Figure EC.1. We have also plotted the combined slope estimates from the main estimations, and comparing against this it can be seen that the direction of the individual effects are consistent with the combined estimates, with 95% confidence intervals overlapping in nearly all cases.

EC.2. Autocorrelated errors

One concern when working with a panel of time series data is that errors may be autocorrelated, i.e. costs change slowly and e.g. high costs in one year may indicate that costs will be high in the next year, also. The standard errors are often underestimated when autocorrelation of the error terms (at low lags) are positive (Hsiao 2015). This is unlikely to be a major issue for our analysis, since results are highly significant and standard errors would have to be vastly underestimated for the results to be misidentified. A bigger concern, however, is that autocorrelation of the errors may bias the coefficient estimates in the within-between model formulation. We investigate this further here.

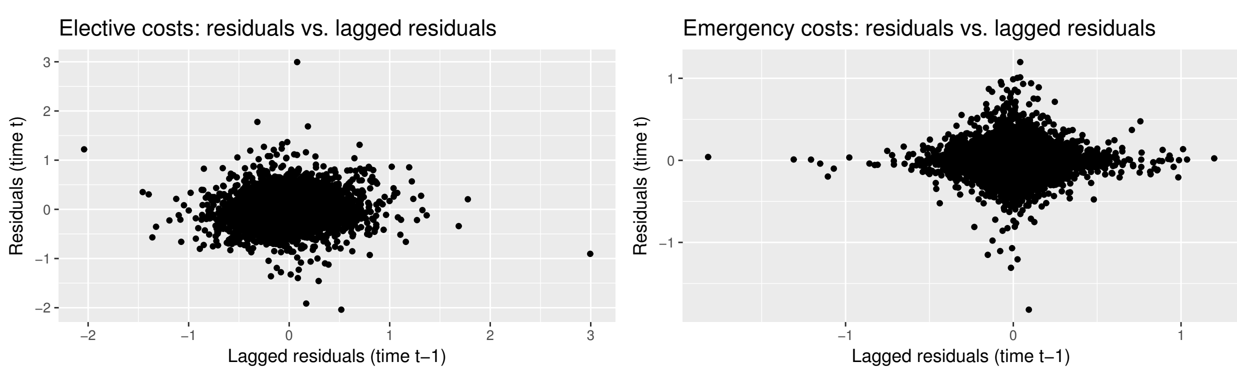
To do this, we have taken three approaches. In the first, we regress (using OLS) the residuals (at time t) from the within-between multilevel models (MLMs) against the lagged residuals (at time $t-1$). A plot of residuals vs. lagged residuals is provided in Figure EC.2, showing little evidence of any correlation and hence suggesting that our models account for much of the within-trust and

Figure EC.1 Random slope coefficient estimates for the effect of volume from the other specialties on the cost of the focal specialty, reported by specialty (black) and combined (red), with bootstrapped 95% confidence intervals.



Note. A - nervous system; B - eyes & periorbita; C - mouth, head, neck, & ears; D - respiratory system; E - cardiac surgery & primary cardiac conditions; F - digestive system; G - hepatobiliary & pancreatic system; H - musculoskeletal system; J - skin, breast & burns; K - endocrine & metabolic system; L - urinary tract & male reproductive system; M - female reproductive system; P - diseases of childhood and neonates; Q - vascular system; V - multiple trauma.

Figure EC.2 Plots of residuals (time t) against lagged residuals (time $t-1$) for elective costs (left) and emergency costs (right).



time-related correlation in the error term. This is confirmed by OLS models, with only $\sim 2.8\%$ of the variance in the residuals for elective costs explained by the lagged residuals, and $< 0.1\%$ for emergency costs.

We follow the informal approach described above with a traditional testing method. The standard test for the presence of first-order correlation is the Durbin-Watson statistic. However, this test can only be performed if the panel is balanced. For an unbalanced panel the recommended approach is to instead calculate the Baltagi-Wu locally best invariant (LBI) test statistic (Baltagi and Wu

1999). We estimate this using the `xtregar` command in Stata 12.1. Note that the models that we estimate this statistic for are not identical to those presented in the paper. This is because the particular command in Stata does not allow the estimation of multiple random effects, and so instead we are only able to include trust–specialty REs. Specifically, we replace Equation (4) in the paper with:

$$\alpha_{(thC)[i]} = \beta^t P_{(t)[i]} + \beta^C P_{(C)[i]} + \beta^h P_{(h)[i]} + \alpha_{(hC)[i]} \,. \tag{EC.1}$$

If anything, since the control structure in the paper includes additional time-related controls (specifically $\alpha_{(th)[i]}$ which has significant explanatory power in the models), the estimates reported here are likely to be conservative. Calculating the LBI statistic we find them to take values 1.70 for elective costs and 1.72 for emergency costs, with estimated AR(1) autocorrelation coefficients equal to 0.28 in both. While critical values are not available in Baltagi and Wu (1999), if there were no evidence of first-order autocorrelation then these should take value 2. While the LBI statistics are close to 2 in value, the fact that the estimated AR(1) coefficients are non-zero indicates that it is worth exploring further.

To extend the above, we re-estimate the models from the paper but where we fit the cross-sectional time-series multilevel models models allowing the disturbance term to be first-order autoregressive. Specifically, if ϵ_{thC} denotes the disturbance term (random error) corresponding to specialties C in hospital trust h at time t, then we can specify that the error term takes the form:

$$\epsilon_{thC} = \rho \times \epsilon_{(t-1)hC} + \xi_{thC} \,. \tag{EC.2}$$

where $|\rho| < 1$ and ξ_{thC} is independent and identically distributed (i.i.d.) with mean 0 and variance σ_z^2. Then ρ estimates the residuals are first-order autoregressive. Estimation is made in R (version 3.3.3) using the `lme()` function of the `nlme` package. One restriction of this package is that implementing non-nested random effects is prohibitively difficult. To get around this, we replace Equation (4) in the paper with:

$$\alpha_{(thC)[i]} = \beta^t P_{(t)[i]} + \beta^C P_{(C)[i]} + \alpha_{(h)[i]} + \alpha_{(hC)[i]} \,. \tag{EC.3}$$

As discussed above, if anything since the control structure in the paper includes additional time-related controls (specifically $\alpha_{(th)[i]}$ which has significant explanatory power in the models), the estimates reported here are likely to overestimate the size of the ρ. We report in Table EC.1 updated coefficient estimates under this new model specification. We observe that all of the results are identical in terms of sign and direction as those reported in Table 2 of the paper, and they are also very similar in terms of scale. Thus, we are confident that the coefficient estimates in the within-between model formulation in the paper are not biased.

Table EC.1 Model parameter estimates – MLMs using within-between volume decomposition and first-order autocorrelated errors

	Costs	
	Elective	Emergency
Longitudinal effects		
Elect. vol. (focal Sp)	-0.108^{***}	0.005
	(0.007)	(0.005)
Emerg. vol. (focal Sp)	0.012	-0.152^{***}
	(0.012)	(0.008)
Elect. vol. (other Sps)	-0.188^{***}	0.056^{***}
	(0.021)	(0.014)
Emerg. vol. (other Sps)	0.014	-0.257^{***}
	(0.020)	(0.014)
Cross-sectional effects		
Elect. vol. (focal Sp)	-0.050^{***}	0.031^{***}
	(0.011)	(0.007)
Emerg. vol. (focal Sp)	−0.008	-0.144^{***}
	(0.018)	(0.011)
Elect. vol. (other Sps)	0.059	0.159^{***}
	(0.032)	(0.028)
Emerg. vol. (other Sps)	−0.058	-0.130^{***}
	(0.036)	(0.031)
Control structure		
Year	Y	Y
Specialty line	Y	Y
Trust	0.086	0.088
Specialty–trust	0.111	0.056
Trust–year	N/A	N/A
Specialty–year	N/A	N/A
Residual std. error	0.242	0.177
Correlation structure: AR(1)		
ρ	0.366^{***}	0.359^{***}
Model fit		
Observations	20,057	21,507
Bayesian inf. crit.	1,192.6	-12,535.7

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; Inclusion of a FE in the control structure indicated by a "Y", inclusion of a RE indicated by the reporting of its estimated standard deviation, else an "N/A" is indicates in the control structure if neither a FE or RE are included.

EC.3. Non-linear volume effects

In the paper we assume the effects of (log) volume on (log) cost is linear, i.e. a 1% increase in volume has an x% effect on cost, regardless of the initial level of volume. Here we discuss relaxing this assumption to allow for non-linear volume effects. We do this by including the squared values of both the longitudinal (within) and cross-section (between–hospital-trust) volume terms in the main multilevel models.

In Figures EC.3 and EC.4 are plotted for the elective and emergency patient types, respectively, the estimated between-effects of volume in models with linear only volume effects (i.e. the estimated effects reported in the paper) and in models with the inclusion of non-linear (squared terms) volume between-effects. 95% confidence bands for the non-linear effects are also plotted. These plots have been restricted to the range over which 98% of the values of the respective volume measures lie

Figure EC.3 Plots of estimated (mean-centered) volume effects on elective costs in models with only linear volume effects (solid lines) and in models also with non-linear volume effects (dashed lines).

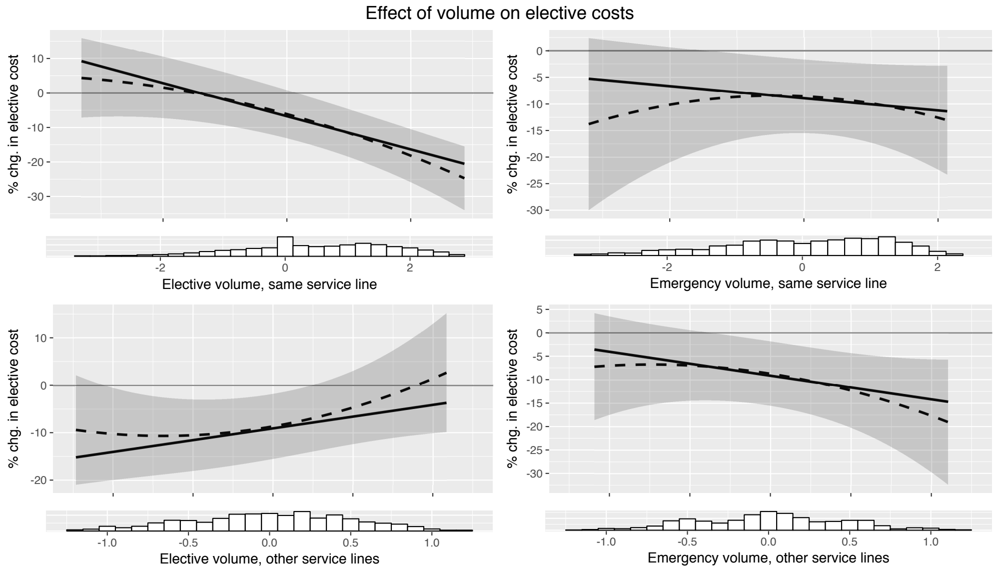
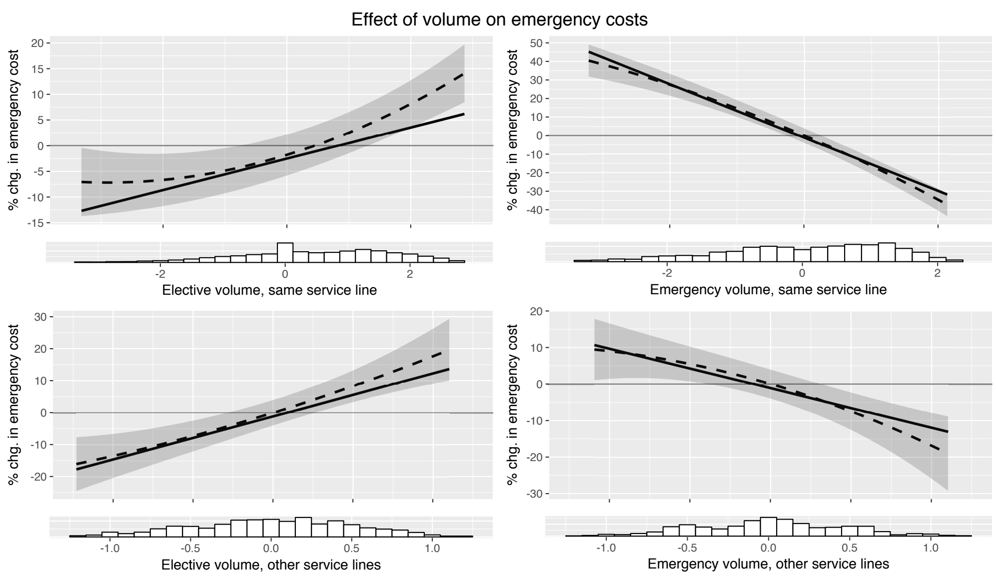


Figure EC.4 Plots of estimated (mean-centered) volume effects on emergency costs in models with only linear volume effects (solid lines) and in models also with non-linear volume effects (dashed lines).



(i.e. excluding the lowest 1% and higher 1%). As shown, there is little evidence to suggest that the interpretation of the results would change significantly if we had instead used a non-linear volume specification.

EC.4. Reverse causality

In Section 5.2 of the paper we discuss three tests that we perform in order to examine whether there is any evidence of reverse causality. Below we provide further details on each of these.

EC.4.1. Cost efficiency driving changes in volume

One possibility, as discussed in the paper, is that the positive relationship identified between volume and productivity is actually the result of more cost effective hospitals being referred or taking action to attract a higher volume of patients. This could be e.g. because patients are referred more often to more productive hospitals or if hospitals use their stronger financial position to take action (e.g. through marketing or lobbying) to increase their patient pool. To test this we run additional analysis where we examine whether past financial performance is a predictor of future patient volumes. If better performing hospitals are able to attract or are referred a higher volume of patients then we would expect lower costs in the past to be positively correlated with higher patient volume in the future.

In order to determine this we closely follow the approach recommended in the multilevel modeling literature (see e.g. Bell et al. 2014). In particular, we specify eight models where we regress future volumes on historic elective (emergency) cost ratios. More specifically, the models are specified as follows:

1. *Dependent variables* – The four dependent variables in these models are set equal to the percentage change between year $t-1$ and year t in the volume of (i) elective activity in the focal specialty, (ii) emergency activity in the focal specialty, (iii) elective activity in all specialty other than the focal specialty, and (iv) emergency activity in all services lines other than the focal specialty.

2. *Primary independent variables of interest* – We use one of two possible independent variables: (a) the standardized cost for elective patients, $CostEl$, in the focal specialty in the previous year, and (b) the standardized cost for emergency patients, $CostEm$, in the focal specialty in the previous year. These are the variables used as the dependent variables in the various models in the paper. An increase in value by one unit at the mean, e.g. from 1 to 2, indicates the cost in that specialty–trust is approximately double that of other trusts.

3. *Controls* – We control for the specialty–year interaction with a fixed effect. This accounts for changes in volume common across all hospitals over the sample period (e.g. due to growth in demand).

If a hospital has lower cost last year relative to other hospitals in a particular specialty then this means that they are likely to have made a profit in that specialty (since our "expected cost" measure used for standardization is approximately equal to the income that a hospital receives). Thus, if lower cost (more profitable) hospitals are able to generate increased demand next year, we should expect to see a lower cost this year translating into an increase in volume next year (i.e. our dependent and independent variables should be negatively correlated). In Table EC.2 we report the direction, size and significance of the estimated coefficients in these eight models. As is shown, when estimating the model described above we find the opposite: the *higher* the cost at a hospital in a particular year, the more likely the hospital is to *increase* activity in the following year.

Table EC.2 The effect of a one unit increase in cost relative to expected cost in the focal specialty on the volume of patients seen by a hospital in the following year.

	Dependent variables: Percentage change in volume between year t and $t-1$			
	Elect. vol (focal Sp)	Emerg. vol (focal Sp)	Elect. vol (other Sps)	Emerg. vol (other Sps)
Elect. cost / exp. cost in year $t-1$	4.61%***	0.55%*	0.95%***	0.17%
Emerg. cost / exp. cost in year $t-1$	-2.55%	11.7%***	0.61%	5.50%***

	Direction of the coefficients in the paper			
Dependent variables (below)	Elect. vol (focal Sp)	Emerg. vol (focal Sp)	Elect. vol (other Sps)	Emerg. vol (other Sps)
Elect. cost / exp. cost	$-$	0	0	0
Emerg. cost / exp. cost	$+$	$-$	$+$	$-$

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.

The effects reported in Table EC.2 suggests insead that higher cost hospitals are likely to respond by trying to increase their activity, perhaps in an effort to increase profitability by exploiting potential economies of scale/increasing utilization. If anything, therefore, in the main paper this is likely to result in higher cost hospitals having *higher* volumes of patients, rather than lower, and work against us finding evidence of economies of scale or positive spillovers. We can also see that there is no evidence that hospitals with high emergency costs in one year attempt to offset those costs by expanding the number of electives that the treat in the next year. In fact, in Table EC.2 we have also reported the direction of the main effects identified in the paper, and the results above suggest that if anything reverse causality is likely to work in the *opposite* direction of all of the main effects that we find.

Further, it is worth pointing out that the influence of cost in one year on volume the next year is very small and unlikely to significantly bias against the results reported in the paper. To see this, suppose that specialty C at a hospital is 10% more costly in treating elective patients than the average hospital in year t (suggesting also that they make a loss of approximately 10% in that year). Then Table EC.2 implies that in year $t+1$ they are likely to expand the volume of electives in that specialty by $10\% \times 4.61\% = 0.46\%$. However, based on the estimated coefficients from Table 2 in the paper, elective volume in the focal specialty would be required to increase by $\sim 210\%$ $(= 0.1/0.048)$ in order to bring about that 10% reduction in cost. Thus, even if the direction of the bias was in the same direction as the estimated coefficients (which it is not), the potential size of the bias is small.

Finally, note that we have also extended the model above to allow cost both in years $t-2$ and $t-1$ to affect costs in year t, and find little evidence of any lagged effect of cost two years prior on volume in the future (results not reported here).

EC.4.2. Patient selection effects

One possibility, as discussed in the paper, is that the positive relationship identified between volume and productivity is actually the result of more patients self-selecting these hospitals. As cost and quality are often correlated, and quality is an unobserved factor that we do not account for in this analysis, this could be driving the results. First note that this seems unlikely to be the case for emergency cases, who almost always attend their nearest hospital, and so we believe that it is appropriate to treat emergency volume as exogenous. However, it is possible that elective patients choose to go to higher quality and hence lower cost (though the link between high quality and low cost is not immediately clear – see below for more on this) hospitals. In the paper we argue that quality information has not been available to patients until recently, but that there may be other more tacit ways of finding out about the quality of a hospital, e.g. by way of word-of-mouth. Below we discuss the test that we perform to look into this further.

To test further whether patients appear to be exercising choice based on quality, we have accessed an "Adult Inpatient Satisfaction Survey" data set (NHS 2017). The survey contains responses to various questions about patient experience at every acute and specialist NHS trust, for which "eligible patients were aged 16 years or over, who had spent at least one night in hospital [...] and were not admitted to maternity or psychiatric units." This data set was first collected during the 2005/06 financial year (before our cost data begins) and has been collected every year since, with the latest data available for the 2015/16 financial year (the last year in our data set). As a result, we are able to match satisfaction scores to 99.7% of the total trust-years in our data set (75 unmatched observations). The survey contains responses from patients to various questions about their inpatient stay, which are aggregated into an "Overall Patient Experience Score". We believe that the overall experience score should thus act as an excellent proxy for "perceived quality", and thus capture much of the "word of mouth" effect that might entice patients to attend certain hospitals over others.

First, it is interesting to look at the correlations between the satisfaction scores and the primary variables in this paper. These are listed below:

- Elective cost: $\rho = -0.027$, p-value < 0.001
- Emergency cost: $\rho = -0.040$, p-value < 0.001
- Elective LOS: $\rho = -0.048$, p-value < 0.001

- Emergency LOS: $\rho = -0.090$, p-value < 0.001
- Cross-sectional elective volume (focal specialty): $\rho = 0.072$, p-value < 0.001
- Cross-sectional emergency volume (focal specialty): $\rho = 0.038$, p-value < 0.001
- Cross-sectional elective volume (other specialties): $\rho = 0.174$, p-value < 0.001
- Cross-sectional emergency volume (other specialties): $\rho = 0.088$, p-value < 0.001

The above correlations suggest that higher quality hospitals (as proxied by greater levels of patient satisfaction) tend to operate at slightly lower cost (the correlations are small but significant) and that they also are able to attract a higher volume of patients (especially elective patients, as we hypothesized above). Note that these statistics are correlations only, and this does not necessary describe a causal relationship, i.e. the higher volume at higher quality hospitals may not only be because patients are attracted to those hospitals, but also because hospitals that operate at higher volume are able to deliver a higher quality of service as has been argued and demonstrated in medical and OM literature.

In order to address whether quality is an important omitted variable, therefore, we have re-run the models from the paper but where the patient satisfaction score is included as an additional control. The satisfaction scores are separated into their longitudinal and cross-sectional components, as per the norm for all of the continuous covariates in the paper. The results after re-estimating the models are presented in Table EC.3.

As is shown in Table EC.3, there is some evidence to suggest that satisfaction scores are higher at hospitals that are able to discharge emergency patients faster, with every one standard deviation increase in the overall satisfaction score resulting in a 3.7% (p-value < 0.001) reduction in emergency LOS and 3.2% (p-value < 0.01) reduction in cost. Note that this may not be causal: instead it could be the case that when a patient is discharged faster they are more likely to report a higher level of satisfaction, rather than the reverse. Regardless, there is no evidence that this has a material impact for the elective patients. This suggests that cost and quality are, for the most part, independent or only weakly dependent (the effect sizes are small when they are significant).

Turning to the coefficients of the four main cross-section volume measure, we see in Table EC.3 that inclusion of this quality metric as a control – in order to capture word-of-mouth effects – results in little to no change in the direction, scale and significance of the coefficient estimates. The only exception is that the effect of emergency volume from the non-focal specialties on emergency LOS in the focal specialty becomes significant at the 5% level (coef. $= -0.052$).

In summary, despite the fact that it is certainly possible that some patients may exercise choice for where they receive elective services, we find no evidence to suggest perceived quality or word-of-mouth effects are an important "omitted variable" that might be driving our results.

Table EC.3 Model parameter estimates – MLMs using within-between volume decomposition with inclusion of overall satisfaction scores as control variables

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Overall satisfaction score	0.002	0.003	−0.005	-0.013^{*}
	(0.005)	(0.005)	(0.003)	(0.005)
Elect. vol. (focal Sp)	-0.130^{***}	0.007	-0.075^{***}	0.005
	(0.006)	(0.004)	(0.003)	(0.003)
Emerg. vol. (focal Sp)	0.002	-0.177^{***}	0.010^{*}	-0.126^{***}
	(0.011)	(0.007)	(0.005)	(0.005)
Elect. vol. (other Sps)	-0.127^{***}	0.088^{***}	-0.030^{*}	0.105^{***}
	(0.028)	(0.026)	(0.014)	(0.024)
Emerg. vol. (other Sps)	0.034	-0.188^{***}	0.029^{*}	-0.111^{***}
	(0.027)	(0.025)	(0.014)	(0.024)
Cross-sectional effects				
Overall satisfaction score	−0.021	-0.032^{**}	−0.006	-0.037^{***}
	(0.012)	(0.010)	(0.005)	(0.009)
Elect. vol. (focal Sp)	-0.047^{***}	0.031^{***}	-0.018^{***}	0.014^{**}
	(0.011)	(0.007)	(0.004)	(0.005)
Emerg. vol. (focal Sp)	−0.012	-0.144^{***}	0.013	-0.107^{***}
	(0.019)	(0.011)	(0.006)	(0.008)
Elect. vol. (other Sps)	0.061	0.151^{***}	0.005	0.068^{**}
	(0.034)	(0.027)	(0.014)	(0.024)
Emerg. vol. (other Sps)	−0.061	-0.125^{***}	0.011	-0.052^{*}
	(0.037)	(0.029)	(0.015)	(0.026)
Model fit				
Observations	19,987	21,432	19,987	21,432
Marginal R^2	0.127	0.222	0.134	0.160
Conditional R^2	0.517	0.624	0.455	0.722
Bayesian inf. crit.	1,797.5	-14,714.5	-26,614.8	-31,014.0

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

EC.4.3. Geographically dispersed hospital trusts

In Table EC.4 we report the within-effects estimated for a subset of hospital trusts constrained to be 20km or more apart (see Section 5.2 of the paper for details). As discussed in the paper, this restriction has the effect of removing those trusts in more urban areas where patients often have more choice as to the provider from which they receive treatment. This effect of this restriction is demonstrated in Figure EC.5, which shows a plot of all trusts (left) together with 20km radius circles, together with a plot of only those that are at least 20km from the nearest alternative trust. Turning to the results in Table EC.4, the main results are comparable in sign and scale to those reported in the paper, though the significant reduction in sample size (a 64% decrease in trust-year observations from 1,434 to 517, and of observations in general from ∼21,507 to ∼7,754) means that standard errors have increased and in some case some cases results no longer appear statistically significant at conventional levels of significance, e.g. the effect of emergency volume from other specialties on emergency costs.

EC.5. Endogenous specialty composition

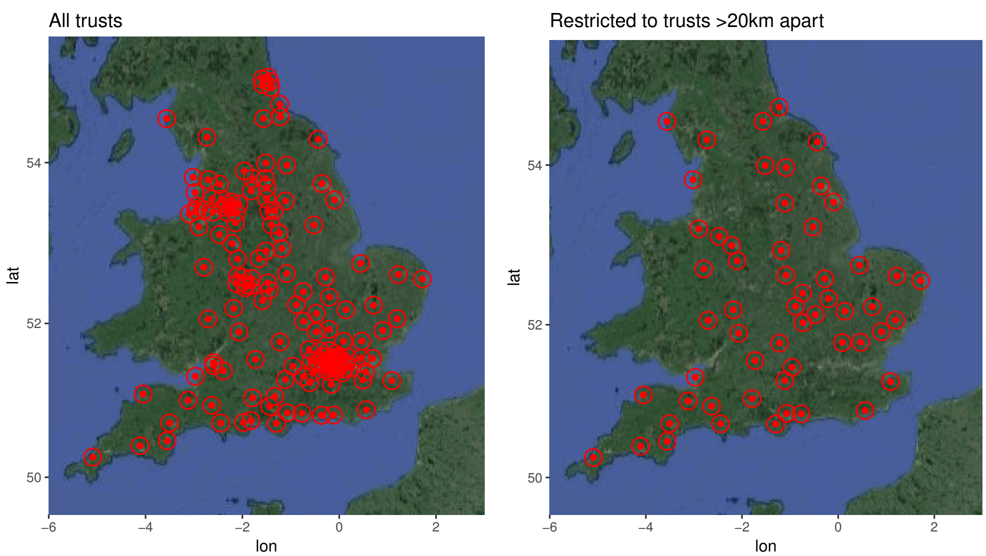
Not every hospital trust may offer every type of treatment, and while hospitals in the UK are not as financially driven as in other healthcare systems, e.g the US, the choice of which treatments to offer

Table EC.4 Model parameter estimates - subset of geographically dispersed hospitals

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.105^{***}	0.023^{**}	-0.071^{***}	0.009
	(0.011)	(0.008)	(0.005)	(0.005)
Emerg. vol. (focal Sp)	-0.018	-0.127^{***}	0.003	-0.097^{***}
	(0.017)	(0.012)	(0.008)	(0.008)
Elect. vol. (other Sps)	-0.117^{*}	0.138^{**}	0.017	0.106^{*}
	(0.052)	(0.051)	(0.024)	(0.042)
Emerg. vol. (other Sps)	0.040	-0.227^{***}	0.035	-0.082^{*}
	(0.047)	(0.046)	(0.021)	(0.039)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.070^{**}	0.020	-0.042^{***}	0.013
	(0.022)	(0.014)	(0.009)	(0.010)
Emerg. vol. (focal Sp)	0.066	-0.126^{***}	0.040^{***}	-0.114^{***}
	(0.037)	(0.024)	(0.012)	(0.017)
Elect. vol. (other Sps)	-0.002	0.123^{*}	0.041	0.121^{*}
	(0.067)	(0.056)	(0.028)	(0.057)
Emerg. vol. (other Sps)	-0.062	-0.102	-0.037	-0.119^{*}
	(0.074)	(0.060)	(0.030)	(0.060)
Model fit				
Observations	7,235	7,754	7,235	7,754
Marginal R^2	0.146	0.171	0.135	0.238
Conditional R^2	0.505	0.567	0.454	0.765
Bayesian inf. crit.	871.1	-3,547.0	-10,082.2	-10,911.2

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.

Figure EC.5 Plots of all trusts (left) and trusts restricted to only those at least 20km furthest from the nearest other trust, with 20km radius circles.



(i.e. the composition of the specialties) might still be related to the financial viability of different treatment options. In the paper we use $\mathbf{Prop}_{thCp} = \sum_{c\in C_{thp}} \alpha_{tcp}$ to control for the extent to which hospital trusts offer either a wide or narrow range of treatment options for particular types of patients or conditions. A plot of these proportions (for each of the specialties) for every trust–year is given in Figures EC.6 and EC.7 for the elective and emergency patient types, respectively.

As can be seen in Figure EC.6 there is some evidence that not all elective treatments are offered

Figure EC.6 Proportion of the "average" elective case-mix offered in each specialty for every trust–year.

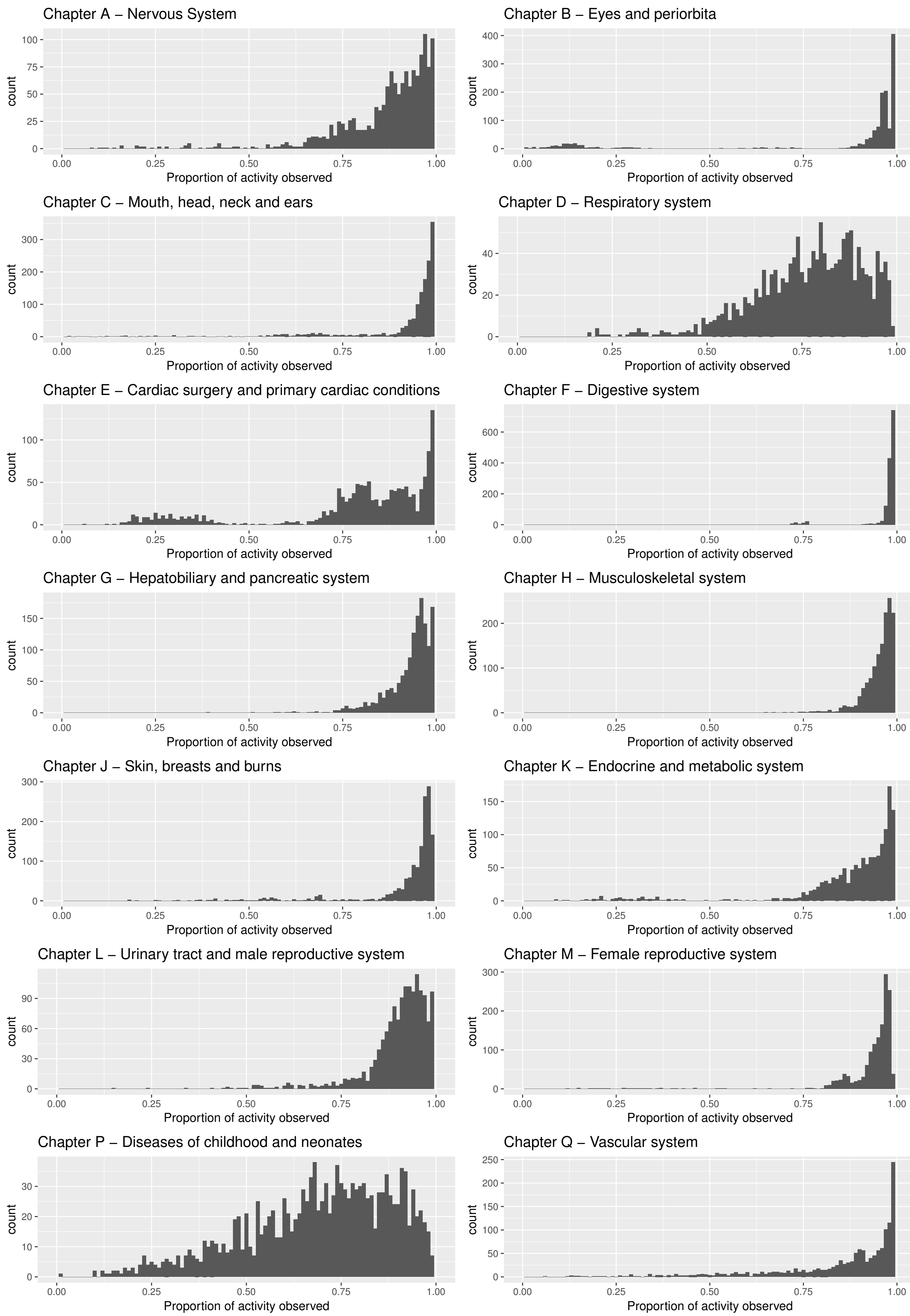
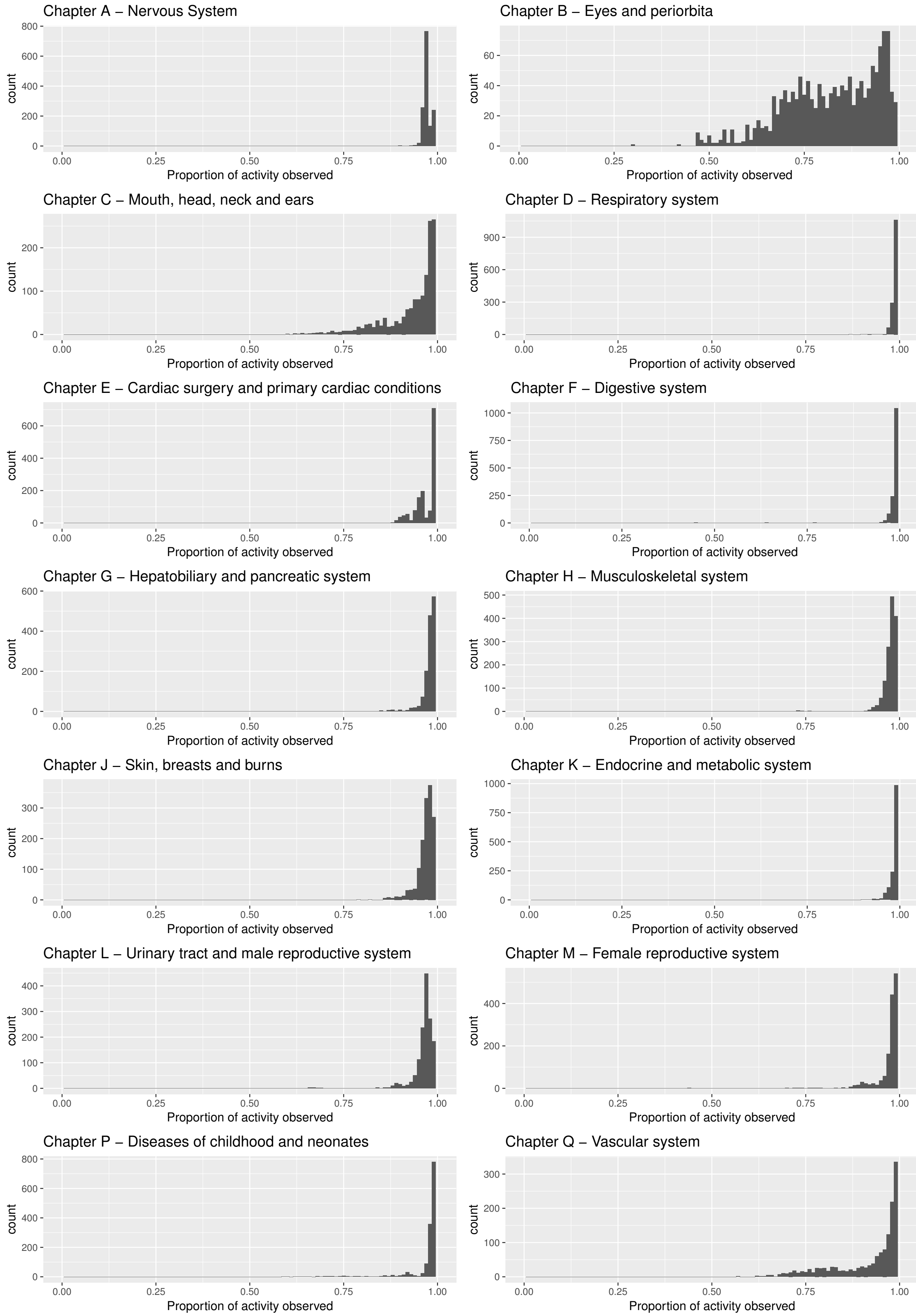


Figure EC.7 Proportion of the "average" emergency case-mix offered in each specialty for every trust–year.



at all hospital trusts, while Figure EC.7 shows that – other than for Chapter B, which relates to conditions of the eyes and periorbita – there is little evidence of emergency treatments not being offered at all trusts (unsurprising, as the unpredictable nature of patient arrivals to the ED means that hospitals have little choice over which emergency patients they serve). For elective specialties, though, it is possible that the mix of services that is offered is formed endogenously, i.e. hospital trusts may choose to only offer treatment to more profitable types of patients. To account for this in the paper we:

1. Construct the dependent variable by dividing actual costs by the 'average' cost, with both calculated using the same weights (i.e. the same case-mix). So, if e.g. only 80% of the HRGs in a specialty appear in the numerator, then only the same 80% of HRGs will appear in the denominator also. In this way costs are adjusted for observable differences in the service offering. More on this can be found in Section 3.1 of the paper under the subheading "Cost Standardization".
2. Use hospital trust and/or trust–specialty fixed- and/or random-effects, to capture systematic, time-invariant differences in the costs at different trusts due to e.g. unobservable differences in the types of treatment offered.
3. Control in the costs and LOS models for $\mathbf{Prop}_{thCp}$ which we have interacted with the specialty C, to capture the fact that costs may differ depending on the range of services offered within a specialty. (In the MLMs we actually control with *both* the longitudinal and cross-sectional values of $\mathbf{Prop}_{thCp}$.)

Despite all of this, we also perform a number of additional tests that we describe in the rest of this Section.

EC.5.1. Relationship between range of services offered, volume and cost

If endogenous formation of the specialty occurred based on cost, then we would expect hospitals that offer a narrow range of services to also be lower cost, since they would opt to only treat patients from profitable HRGs. To determine this, let pEl and pEm specify $\mathbf{Prop}_{thCp}$ when patient admission category $p = El$ and $p = Em$ respectively, with pEl^{CS} and pEm^{CS} the corresponding cross-section values, and $pEl^{LT} = pEl - pEl^{CS}$ and $pEm^{LT} = pEm - pEm^{CS}$ the corresponding longitudinal values. Then we can check whether hospitals that offer a narrow range of services are lower cost by observing the coefficient estimates of pEl^{CS}, pEm^{CS}, pEl^{LT} and pEm^{LT}.

Ideally we would report the above coefficients directly from the model in the paper. However, one problem with this is that, as noted in Section 3.4, in the paper we interact $\mathbf{Prop}_{thCp}$ by the specialty fixed effect, P_C. The problem with this is that we therefore do not estimate the *global* effects of pEl^{CS}, pEm^{CS}, pEl^{LT} and pEm^{LT}, instead we estimate a specific effect for each specialty.

Table EC.5 Model parameter estimates for $propEl$ and $propEm$ – MLMs using within-between volume decomposition and random service line dependent slopes for $propEl$ and $propEm$

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Prop. elect.	0.148	0.003	0.171^{***}	0.010
	(0.100)	(0.025)	(0.029)	(0.022)
Prop. emerg.	−0.010	−0.002	0.053	−0.180
	(0.107)	(0.125)	(0.093)	(0.138)
Elect. vol. (focal Sp)	-0.130^{***}	$0.007^{\dagger}$	-0.074^{***}	0.004
	(0.006)	(0.004)	(0.003)	(0.003)
Emerg. vol. (focal Sp)	0.002	-0.176^{***}	$0.010^{\dagger}$	-0.126^{***}
	(0.011)	(0.007)	(0.006)	(0.005)
Elect. vol. (other Sps)	-0.126^{***}	0.086^{***}	-0.031^{*}	0.100^{***}
	(0.028)	(0.026)	(0.014)	(0.023)
Emerg. vol. (other Sps)	0.037	-0.185^{***}	0.028^{*}	-0.107^{***}
	(0.027)	(0.025)	(0.014)	(0.024)
Cross-sectional effects				
Prop. elect.	0.170	-0.123^{**}	−0.025	$-0.074^{\dagger}$
	(0.112)	(0.042)	(0.051)	(0.040)
Prop. emerg.	$0.446^{\dagger}$	1.098^{***}	$0.277^{\dagger}$	0.291^{*}
	(0.237)	(0.171)	(0.145)	(0.127)
Elect. vol. (focal Sp)	-0.045^{***}	0.031^{***}	-0.021^{***}	0.014^{**}
	(0.011)	(0.006)	(0.004)	(0.005)
Emerg. vol. (focal Sp)	−0.018	-0.139^{***}	0.011	-0.105^{***}
	(0.018)	(0.011)	(0.008)	(0.008)
Elect. vol. (other Sps)	0.041	0.133^{***}	0.003	0.049^{*}
	(0.033)	(0.027)	(0.013)	(0.024)
Emerg. vol. (other Sps)	−0.035	-0.110^{***}	0.013	−0.035
	(0.036)	(0.029)	(0.015)	(0.026)
Observations	20,057	21,507	20,057	21,507

$^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

We would like, therefore, to estimate the *global* effect but also allow for the fact that there may be specialty-specific differential effects. To achieve this, we adopt an approach similar to that used for volume in Section 5.1 of the paper and described further in EC.1 of this document, i.e. we estimate a *global* effect plus random slopes for each of the proportion measures.[8]

Using the random slopes model described above, we present the estimated coefficients in Table EC.5. The coefficients of interest – for pEl^{CS}, pEm^{CS}, pEl^{LT} and pEm^{LT} – are given in the rows denoted Prop. elect. and Prop. emerg.

As is shown in Table EC.5, for emergencies, the greater the proportion of services offered within a particular specialty (i.e. the wider the scope and more varied the service offering), the greater also the cost (coef. 1.098). For electives. the effect is positive though insignificant (coef. 0.170). This indicates, as hypothesized above, that those hospital trusts that choose to offer a narrow range of services tend also to operate at lower cost, perhaps by only providing those less costly services that they are able to deliver more efficiently.

However, if we look at the correlation between pEl^{CS} and $nElS^{CS}$, and between pEm^{CS} and

[8] Since the model is slightly different from the one used in the paper, we will also need to show that the main volume effects remain consistent. Therefore we also report the estimated coefficients of the volume measures in Table EC.5 – all of which are consistent in terms of sign, scale and significance with those in the paper.

$nEmS^{CS}$ we find it to be positive and highly significant, taking values 0.78 and 0.61 respectively. This suggests that higher volume hospitals are less selective in their service offering (i.e. they offer a wider range of services and so $\mathbf{Prop}_{thCp}$ is higher). Since from Table EC.5 we see that hospitals that offer a wider range of services tend to be more expensive, as a consequence we would therefore expect larger hospitals to be more expensive as they do not selectively choose cheaper services to offer. This would work against the findings in our paper (i.e. we find evidence of economies of scale), suggesting that endogenous formation of service offerings within a specialty is not driving our results.

EC.5.2. Diversified hospitals

To confirm the robustness of the results, we have re-run the analysis in the paper for a subset of the data in which we only include observations corresponding to (i) specialty–trusts for which $\mathbf{Prop}_{thCp} > 0.95$ for $p \in \{El, Em\}$ in at least 50% of the years they are represented in the data, i.e. the hospital trust treats at least 95% of the expected case-mix in that specialty in at least 50% of the years, and (ii) specialty–trust–years for which $\mathbf{Prop}_{thCp} > 0.95$ for $p \in \{El, Em\}$, i.e. dropping all observations where less than 95% of the expected case-mix was treated. This reduces the sample by 57.0% for the elective patient type, and 18.3% for the emergency patient type. In these models we do not control for $p \in \{El, Em\}$. This alleviates concerns that the results in the paper may be spurious and caused by the high correlation between the proportion of conditions treated and the volume measures. The findings are reported in Table EC.6.

The results in Table EC.6 are consistent with those documented in the paper, with the exception that we now also see that an increase in the volume of emergencies within the focal specialty results in an increase in the cost of the electives. This suggests that if we restrict our analysis to the subsample of specialty–trusts that offer the full spectrum of services in most of the years, then any increase in within–specialty volume of one admission type may drive up the costs of patients of the other admission type. While we do not report this report in the paper, since it is estimated from a more limited subsample of the data, we note that this finding does not counteract our finding that there are negative spillovers between admission types within a specialty, it only extends it.

Overall, although endogenous specialty formation is a valid concern, we have demonstrated that it is extremely unlikely to be driving the results reported in the paper.

EC.6. Modeling – data alternatives

In this section we report the results from a number of other estimations made using:

Table EC.6 Model parameter estimates - MLMs where observations for which a low proportion of the case-mix is treated are excluded

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.178^{***}	0.009^{*}	-0.051^{***}	0.001
	(0.011)	(0.004)	(0.005)	(0.003)
Emerg. vol. (focal Sp)	0.035^{**}	-0.229^{***}	-0.003	-0.180^{***}
	(0.013)	(0.008)	(0.006)	(0.005)
Elect. vol. (other Sps)	-0.085^{**}	0.096^{***}	-0.007	0.108^{***}
	(0.030)	(0.025)	(0.014)	(0.023)
Emerg. vol. (other Sps)	0.008	-0.144^{***}	0.039^{**}	-0.091^{***}
	(0.029)	(0.024)	(0.013)	(0.024)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.086^{***}	0.036^{***}	-0.016^{*}	0.020^{***}
	(0.015)	(0.005)	(0.007)	(0.004)
Emerg. vol. (focal Sp)	0.062^{**}	-0.147^{***}	0.024^{**}	-0.119^{***}
	(0.020)	(0.011)	(0.009)	(0.008)
Elect. vol. (other Sps)	0.046	0.100^{***}	0.003	0.034
	(0.031)	(0.026)	(0.013)	(0.024)
Emerg. vol. (other Sps)	-0.065	-0.077^{**}	0.001	-0.019
	(0.037)	(0.029)	(0.015)	(0.026)
Model fit				
Observations	8,627	17,576	8,627	17,576
Marginal R^2	0.158	0.201	0.193	0.148
Conditional R^2	0.619	0.667	0.661	0.787
Bayesian inf. crit.	-4,231.7	-16,211.6	-18,675.2	-29,431.2

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

EC.6.1. Capped costs

It is not uncommon for hospital trusts' costs to be magnified (or shrunk) by a few extremely expensive (or low-cost) patients. Therefore, when government agencies calculate hospital trust compensation based on HRG tariffs, the costs are often trimmed to exclude extreme observations. We adopt a similar approach by limiting the influence of "extreme" costs by capping them at a minimum or a maximum value. We do this by constraining in every hospital trust h the average cost of treating patients with HRG c and of patient type p (i.e. cost_{thcp}) to take maximum value equal to 5 multiplied by the across-trust median in that year t, and minimum value equal to 1/5 multiplied by the across-trust median. These caps leave the same sample as in the paper, but limits the extent to which extreme values for individual cost can affect the results.

We report in Table EC.7 the results of the cost and LOS estimations, which are nearly identical to those reported in the paper.

EC.6.2. Common HRGs

In the paper we compared costs and LOS at hospital trusts across the set of all HRGs c treated in year t. An alternative to this would have been to compare hospital trusts across a set of common HRGs, i.e. excluding those conditions that are more rare for which treatment is typically only offered in large, teaching or specialist hospitals. This has the additional benefit that it partially reduces the potential for bias caused by endogeneous specialty formation (see Section EC.5 of this

Table EC.7 Model parameter estimates – MLMs using within-between volume decomposition

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.123^{***}	0.006	-0.075^{***}	$0.005^{\dagger}$
	(0.006)	(0.004)	(0.003)	(0.003)
Emerg. vol. (focal Sp)	0.001	-0.171^{***}	0.010^{*}	-0.129^{***}
	(0.010)	(0.007)	(0.005)	(0.005)
Elect. vol. (other Sps)	-0.124^{***}	0.082^{***}	-0.032^{*}	0.101^{***}
	(0.026)	(0.024)	(0.014)	(0.024)
Emerg. vol. (other Sps)	0.032	-0.174^{***}	0.030^{*}	-0.106^{***}
	(0.026)	(0.024)	(0.014)	(0.024)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.047^{***}	0.026^{***}	-0.018^{***}	0.013^{**}
	(0.010)	(0.006)	(0.004)	(0.005)
Emerg. vol. (focal Sp)	-0.009	-0.137^{***}	0.015^{**}	-0.107^{***}
	(0.018)	(0.010)	(0.006)	(0.008)
Elect. vol. (other Sps)	0.045	0.123^{***}	0.001	$0.046^{\dagger}$
	(0.031)	(0.026)	(0.013)	(0.024)
Emerg. vol. (other Sps)	-0.049	-0.098^{***}	0.014	-0.031
	(0.034)	(0.028)	(0.014)	(0.026)
Model fit				
Observations	20,057	21,507	20,057	21,507
Marginal R^2	0.130	0.217	0.135	0.146
Conditional R^2	0.528	0.640	0.457	0.724
Bayesian inf. crit.	-986.9	-18,287.1	-26,792.7	-31,134.2

$^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

document), since we only compare costs against a base set of HRGs c that are widely provided (and so typically higher volume also, with less discretion in their provision).

To achieve this, we specify that an HRG is only included in the comparison if it is provided by at least 80% of baseline set, T_b, of 116 reference trusts in a particular year. (See Section 3.1 of the paper for more on the baseline trusts.) In the paper we compare hospital costs across approximately 1,500 elective HRGs and 1,400 emergency HRGs per year on average. Once we apply the above condition, we instead are comparing hospital costs across approximately 480 elective HRGs and 750 emergency HRGs per year on average. While we compare across significantly fewer HRGs, we note that these capture 87.8% (78.5%) of the total elective activity (cost) and 96.6% (91.1%) of the total emergency activity (cost) over the sample period, respectively. This indicates clearly that those HRGs that are kept in the sample are those that are higher volume and more prevalent across hospitals. Importantly for the analysis presented here, the volume metrics are left unchanged and are equal to the four original volume measures used in the paper. Note also that the sample size goes down slightly since there are are some hospitals that by random chance only treat rarer cases of certain conditions within a specialty. As such, once we restrict the sample to only the most common set of conditions we may have no cost or LOS data associated with some specialty–trust–years.

The results, using the same estimation method as in the paper, are provided in Table EC.8. As can be seen, the sign, direction and significance of the estimation on the subset of more common HRGs are similar to those reported in the paper.

Table EC.8 Model parameter estimates - calculated on a set of common HRGs

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.142^{***}	0.012^{**}	-0.069^{***}	0.006^{*}
	(0.007)	(0.004)	(0.003)	(0.003)
Emerg. vol. (focal Sp)	0.015	-0.158^{***}	0.008	-0.120^{***}
	(0.012)	(0.007)	(0.006)	(0.005)
Elect. vol. (other Sps)	-0.127^{***}	0.087^{***}	-0.034^{*}	0.099^{***}
	(0.029)	(0.026)	(0.014)	(0.024)
Emerg. vol. (other Sps)	0.024	-0.209^{***}	0.029^{*}	-0.115^{***}
	(0.029)	(0.026)	(0.014)	(0.024)
Cross-sectional effects				
Elect. vol. (focal Sp)	$-0.017^{\dagger}$	0.032^{***}	-0.014^{***}	0.013^{**}
	(0.010)	(0.006)	(0.004)	(0.004)
Emerg. vol. (focal Sp)	-0.025	-0.061^{***}	0.015^{*}	-0.069^{***}
	(0.017)	(0.010)	(0.006)	(0.007)
Elect. vol. (other Sps)	0.025	0.109^{***}	0.001	0.030
	(0.034)	(0.028)	(0.013)	(0.024)
Emerg. vol. (other Sps)	-0.012	-0.136^{***}	0.019	-0.041
	(0.037)	(0.030)	(0.014)	(0.026)
Model fit				
Observations	20,021	21,471	20,021	21,471
Marginal R^2	0.111	0.184	0.125	0.129
Conditional R^2	0.498	0.624	0.424	0.716
Bayesian inf. crit.	4,960.7	-14,818.1	-24,639.5	-30,425.0

$^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

EC.6.3. Minimum specialty size

In Section EC.6.2 above we consider the possibility that the composition of HRGs used to compare hospital trusts (in particular, the inclusion of rarer conditions) may affect the results. Another possibility is that the inclusion of hospital trusts which treat only a low volume of activity within a particular specialty (i.e. which provide no or only a limited scope of service) may be outliers and may be influencing the results.

To examine this, we have re-run the cost models from the paper on a subset of the data such that only those years in which a trust treats at least 25% of the median elective volume and emergency volume of activity within a particular specialty are included in the sample. The median is calculated in each year across the baseline set, T_b, of 116 reference trusts. (See Section 3.1 of the paper for more on the baseline trusts.) This reduces the sample by ∼7% (∼2%) from 20,057 (21,507) elective (emergency) observation to 18,701 (21,171). The results – provided in Table EC.9 – are almost identical to those in the paper, suggesting the findings are not heavily influenced by the presence of trust–specialty with a low volume of activity.

EC.6.4. Multi-site versus single-site hospitals

The analysis in the paper was run on the set of all trusts operating in England. As mentioned in Section 3.4 of the paper, trusts may operate multiple hospitals across multiple sites. While often there is a main hospital site that treats the vast majority of the patients, there are a number of hospital trusts (e.g. Guy's and St. Thomas' in London) where the same trust operates multiple large

Table EC.9 Model parameter estimates - excluding hospital-years with low service line volume

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.158^{***}	0.010^{*}	-0.075^{***}	0.005
	(0.008)	(0.004)	(0.004)	(0.003)
Emerg. vol. (focal Sp)	0.013	-0.202^{***}	0.010^{*}	-0.142^{***}
	(0.011)	(0.007)	(0.005)	(0.005)
Elect. vol. (other Sps)	-0.092^{***}	0.084^{***}	-0.016	0.094^{***}
	(0.027)	(0.026)	(0.014)	(0.023)
Emerg. vol. (other Sps)	0.016	-0.173^{***}	0.029^{*}	-0.102^{***}
	(0.027)	(0.025)	(0.013)	(0.024)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.054^{***}	0.030^{***}	-0.020^{***}	0.015^{**}
	(0.012)	(0.007)	(0.005)	(0.005)
Emerg. vol. (focal Sp)	-0.008	-0.148^{***}	0.017^{**}	-0.109^{***}
	(0.019)	(0.011)	(0.006)	(0.008)
Elect. vol. (other Sps)	0.041	0.136^{***}	0.005	0.048^{*}
	(0.032)	(0.027)	(0.013)	(0.024)
Emerg. vol. (other Sps)	-0.049	-0.105^{***}	0.009	-0.035
	(0.036)	(0.029)	(0.015)	(0.026)
Model fit				
Observations	18,701	21,171	18,701	21,171
Marginal R^2	0.120	0.216	0.123	0.146
Conditional R^2	0.542	0.633	0.509	0.734
Bayesian inf. crit.	-1,418.3	-15,250.7	-29,468.8	-31,557.7

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

hospitals. As it is not possible in our data to distinguish between which patients were treated at which site, it could for such hospital trusts be the case that specialties and/or elective or emergency patients are split over multiple sites. The scale and scope effects we identify may be affected by this, despite the fact that we have taken steps to account for this with control variables.

EC.6.4.1. Single hospital trusts To investigate this further, we have repeated the analysis from the paper using a subset of the data corresponding to those trusts that only operate a single hospital site. This has the effect of reducing the sample by 43.3%, from 21,507 observations to 12,192. In these models we remove the controls for (i) the number of sites operated by the hospital, and (ii) the concentration of beds across sites, since there are equal for all single hospital sites. The results, reported in Table EC.10, show that even when restricting the sample there is little change in the sign or scale of the main results reported in the paper. As such, in the paper we report the results from all trusts.

EC.6.4.2. Concentrated trusts We also repeat the analysis of the paper for the subset of trusts where either (i) the hospital operates only a single hospital site, or (ii) the hospital operates multiple hospital sites but the beds are highly concentrated in a single site. In particular, to satisfy (ii) we require that at least 80% of the beds operated by that hospital trust are located in a single site. Applying this restriction reduces the sample to 14,442 observations (a reduction of approx. 32.8%). Results, reported in Table EC.11, show again that there is no evidence of a change to our findings when applying this restriction.

Table EC.10 Model parameter estimates - subset of trusts operating one hospital site

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.100^{***}	0.014^{*}	-0.073^{***}	0.011^{**}
	(0.008)	(0.006)	(0.004)	(0.004)
Emerg. vol. (focal Sp)	-0.004	-0.158^{***}	0.011	-0.107^{***}
	(0.015)	(0.009)	(0.007)	(0.007)
Elect. vol. (other Sps)	-0.156^{***}	0.086^{*}	-0.026	0.153^{***}
	(0.035)	(0.034)	(0.018)	(0.032)
Emerg. vol. (other Sps)	0.096^{**}	-0.243^{***}	$0.035^{\dagger}$	-0.155^{***}
	(0.035)	(0.034)	(0.018)	(0.032)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.046^{**}	0.027^{**}	-0.017^{**}	0.013^{*}
	(0.014)	(0.009)	(0.006)	(0.006)
Emerg. vol. (focal Sp)	-0.004	-0.129^{***}	0.016^{*}	-0.109^{***}
	(0.025)	(0.015)	(0.008)	(0.011)
Elect. vol. (other Sps)	0.046	0.113^{***}	0.001	$0.061^{\dagger}$
	(0.038)	(0.034)	(0.015)	(0.032)
Emerg. vol. (other Sps)	-0.044	-0.095^{*}	0.022	-0.048
	(0.046)	(0.040)	(0.018)	(0.037)
Model fit				
Observations	11,363	12,192	11,363	12,192
Marginal R^2	0.134	0.216	0.147	0.148
Conditional R^2	0.496	0.622	0.429	0.718
Bayesian inf. crit.	2,821.1	-6,339.4	-12,773.6	-14,679.7

$^{\dagger}p<0.10$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.

Table EC.11 Model parameter estimates - subset of trusts operating one or more sites where beds highly concentrated in single site

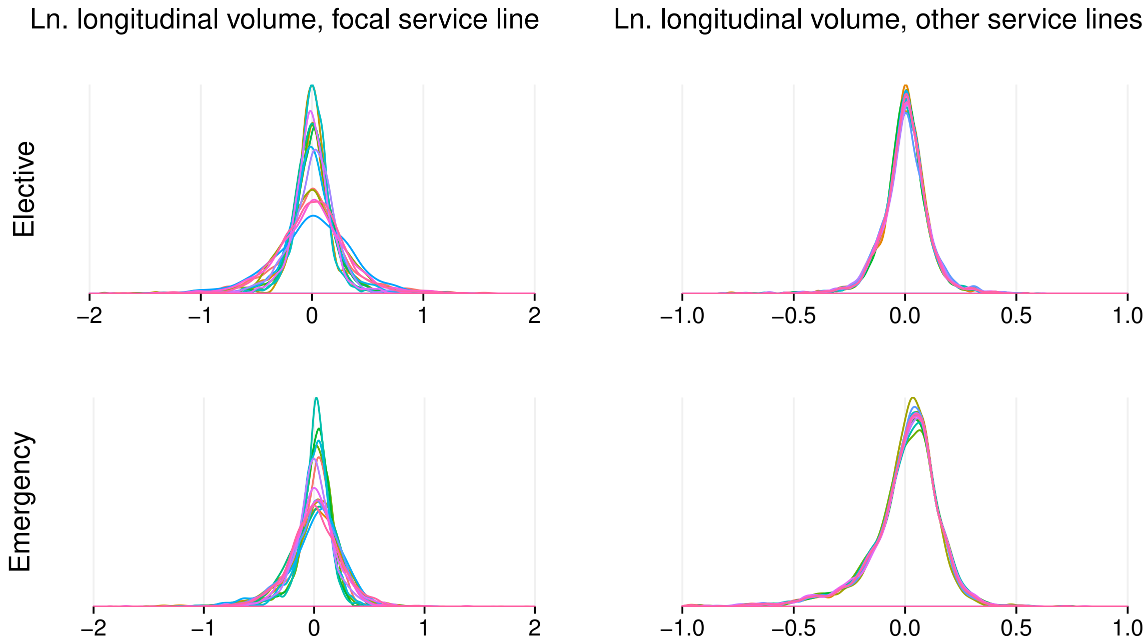
	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.126^{***}	0.014^{**}	-0.076^{***}	0.009^{*}
	(0.008)	(0.005)	(0.004)	(0.004)
Emerg. vol. (focal Sp)	0.003	-0.169^{***}	$0.011^{\dagger}$	-0.115^{***}
	(0.014)	(0.009)	(0.006)	(0.006)
Elect. vol. (other Sps)	-0.138^{***}	$0.061^{\dagger}$	-0.015	0.138^{***}
	(0.035)	(0.032)	(0.018)	(0.031)
Emerg. vol. (other Sps)	0.092^{**}	-0.208^{***}	0.026	-0.148^{***}
	(0.033)	(0.031)	(0.017)	(0.031)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.052^{***}	0.032^{***}	-0.015^{**}	$0.011^{\dagger}$
	(0.013)	(0.008)	(0.005)	(0.006)
Emerg. vol. (focal Sp)	-0.011	-0.148^{***}	0.010	-0.118^{***}
	(0.023)	(0.013)	(0.007)	(0.010)
Elect. vol. (other Sps)	0.056	0.137^{***}	-0.005	$0.048^{\dagger}$
	(0.039)	(0.032)	(0.014)	(0.029)
Emerg. vol. (other Sps)	-0.019	-0.094^{*}	0.021	-0.046
	(0.047)	(0.037)	(0.016)	(0.033)
Model fit				
Observations	13,463	14,442	13,463	14,442
Marginal R^2	0.120	0.212	0.134	0.158
Conditional R^2	0.509	0.626	0.426	0.728
Bayesian inf. crit.	2,945.9	-8,309.8	-15,502.9	-18,259.6

$^{\dagger}p<0.10$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.

EC.6.5. Asset changes

Since the panel spans 10 years, another concern might be that the asset structure of the hospitals changes over that period. If this were the case then the assumption of fixed capacity as noted in Footnote 4 would be violated. This might influence the cost and volumes at a hospital

Figure EC.8 Distribution of longitudinal (within hospital) volume by specialty: natural logarithm of focal specialty volume (left) and other specialties volume (right), for elective (top) and emergency (bottom) admissions.



simultaneously, and may lead to spurious results. First, note that we account for any time-varying variation in cost common to all specialties within a hospital with the trust–year controls. Thus, if time-varying unobserved heterogeneity affects our results then this must occur at the specialty level within an individual hospital. We delve into this below.

In Figure EC.8 we have plotted the distribution of specialty–trust level longitudinal volume (i.e. the difference between volume in a particular year and the mean volume across all years) for each specialty. This is on the log scale, indicating that although rare, there are some instances where specialties exhibit reasonably large changes in volume that may be worth exploring further (note that this is more likely to occur at the level of an individual specialty, than at a hospital as a whole, demonstrated by the greater variation for the focal specialty shown in the left-hand column of Figure EC.8).

In order to check whether there is any evidence that our results are affected by potential structural changes within a hospital/specialty over time, we have repeated our analysis but have split the time horizon in two, so that the maximum period over which we assume the asset configuration at a hospital remains relatively stable is 5 years, rather than 10. To achieve this we do the following:

- If a hospital trust h is observed in the sample for 6 or more years, we separate the observations for that trust into two. Specifically, if t_h is the number of years that hospital h is observed, we separate the observations corresponding to the first floor$(t_h/2)$ years and last ceiling$(t_h/2)$ years,

Table EC.12 Model parameter estimates – Split sample

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.107^{***}	0.003	-0.071^{***}	0.002
	(0.008)	(0.005)	(0.004)	(0.003)
Elect. vol. (other Sps)	-0.110^{**}	0.020	$-0.033^{\dagger}$	0.076^{*}
	(0.036)	(0.034)	(0.018)	(0.031)
Emerg. vol. (focal Sp)	0.004	-0.152^{***}	0.006	-0.114^{***}
	(0.013)	(0.008)	(0.007)	(0.005)
Emerg. vol. (other Sps)	-0.002	-0.183^{***}	0.023	-0.081^{**}
	(0.029)	(0.027)	(0.015)	(0.026)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.065^{***}	0.029^{***}	-0.033^{***}	0.012^{***}
	(0.008)	(0.005)	(0.004)	(0.004)
Elect. vol. (other Sps)	0.024	0.133^{***}	0.0003	0.055^{**}
	(0.026)	(0.022)	(0.011)	(0.019)
Emerg. vol. (focal Sp)	-0.011	-0.145^{***}	0.015^{*}	-0.103^{***}
	(0.014)	(0.008)	(0.006)	(0.006)
Emerg. vol. (other Sps)	-0.012	-0.103^{***}	0.012	-0.041^{*}
	(0.029)	(0.024)	(0.012)	(0.021)
Model fit				
Observations	20,057	21,507	20,057	21,507
Marginal R^2	0.119	0.211	0.135	0.136
Conditional R^2	0.575	0.673	0.506	0.755
Bayesian inf. crit.	1,399.9	-15,266.4	-26,395.6	-31,348.6

$^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

and treat these as belonging to two separate organization (i.e. we generate two new trust indicators h_1 and h_2, corresponding to the two periods). This increases the effective number of trusts from 169 to 312.

- We re-generate cross-section and longitudinal volume measures for the new set of 312 trusts.
- We re-run the multilevel models with the updated volume measures.

Another way to think of this is that we effectively allow each hospital to have one major structural midway through the observation period (so long as they are observed for 6 or more years). This will act to capture some of the potential time-varying heterogeneity. The estimated model coefficients under this updated specification are supplied in Table EC.12 below.

As can be seen, all of the results from the paper continue to hold even when we allow for structural changes in hospitals over the sample period. In fact, comparing to the main results in the paper shows there is very little change in estimated coefficients. This suggests that time-varying unobserved heterogeneity at the specialty level (not trust level, since this is already captured with trust-year level random effects) is unlikely to be an important omitted component of our control structure. To see why, note that in the models we run above we effectively double the number of specialty–trust random effects. These additional random effects should capture a component of any potential time-varying unobserved heterogeneity, since the random effects are allowed to take different values across the two periods for each trust and specialty (with volume differences already picked up with the updated cross-sectional volume measures).

Given the fact that our results do not change after accounting for time-varying unobserved heterogeneity as described above, we have little reason to be concerned that this plays a significant role here.

EC.7. Combined panel analysis

In the main paper we perform separate analysis for the subset of elective costs and emergency costs. There we two reasons for doing this: (1) we have no reason to believe apriori that the impact of each of the covariates (both controls and the volume effects of interest) on costs will be the same for emergencies and electives, and (2) in order to control properly in this model would require us to go from a three dimensional panel (year, trust, specialty) to four dimensional (year, trust, specialty, patient type), and this significantly increases the number of random and fixed effects that must be estimated in the model, and hence the computation time. While point (1) can be resolved by interacting the independent variables with the patient type, point (2) is more problematic, especially given the large number of robustness checks required in order to ensure the validity and reliability of the results. One problem with the approach in the paper, however, is that it inherently assumes that the errors across the two panels (electives and emergencies) are uncorrelated. There may be reason to suspect that this should not be the case, though, since e.g. if the cost of elective patients within a particular specialty at a particular hospital is high (or low) this may suggest that the cost of emergency patients within the same specialty and hospital trust will also be high (or low).

To test whether our results are robust to re-specification where we allow elective and emergency costs to be correlated, we have re-estimated the main results presented from the paper under a new model specification given as follows:

$$\begin{aligned} ln(Cost_i) = {} & \alpha_{(thCp)[i]} + (\beta_1^{LT} nElH_i^{LT} + \beta_2^{LT} nElS_i^{LT} + \beta_3^{LT} nEmH_i^{LT} + \beta_3^{LT} nEmS_i^{LT} \\ & + \beta_1^{CS} nElH_i^{CS} + \beta_2^{CS} nElS_i^{CS} + \beta_3^{CS} nEmH_i^{CS} + \beta_3^{CS} nEmS_i^{CS}) : Type + \epsilon_i \,, \end{aligned} \tag{EC.4}$$

where $: Type$ denotes an interaction between the volume effects and the patient type (elective or emergency), and the intercept is given by

$$\begin{aligned} \alpha_{(thCp)[i]} = {} & \mathbf{bX} + \beta^t P_{(t)[i]} + \beta^C P_{(C)[i]} + P_{(p)[i]} + \alpha_{(h)[i]} + \alpha_{(th)[i]} + \alpha_{(tC)[i]} + \alpha_{(hC)[i]} \\ & + \alpha_{(hp)[i]} + \alpha_{(thp)[i]} + \alpha_{(tCp)[i]} + \alpha_{(hCp)[i]} \,. \end{aligned} \tag{EC.5}$$

Using the notation recommended in Gelman and Hill (2007), the index $(thCp)[i]$ denotes the time, t, hospital trust, h, specialty, C, and patient type p, corresponding to observation i, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

is the idiosyncratic error term. The terms $\alpha_{(x)[i]}$, where $(x)[i]$ takes values $(h)[i]$, $(th)[i]$, $(tC)[i]$, $(hC)[i]$, $(hp)[i]$, $(th)[i]$, $(tCp)[i]$, and $(hCp)[i]$, denote the hospital trust, trust–year, specialty–year and specialty–trust, trust–patient-type, trust–year–patient-type, specialty–year–patient-type and specialty–trust–patient-type random effects (REs), respectively, which are all assumed to be Normal random variables with a standard deviation to be estimated.

Table EC.13 Model parameter estimates – MLMs using within-between volume decomposition

	Costs		LOS	
	Elective	Emergency	Elective	Emergency
Longitudinal effects				
Elect. vol. (focal Sp)	-0.130^{***}	0.006	-0.074^{***}	0.005
	(0.005)	(0.005)	(0.003)	(0.003)
Emerg. vol. (focal Sp)	0.002	-0.180^{***}	0.014^{**}	-0.130^{***}
	(0.009)	(0.009)	(0.005)	(0.005)
Elect. vol. (other Sps)	-0.109^{***}	0.087^{***}	-0.035^{*}	0.104^{***}
	(0.027)	(0.027)	(0.020)	(0.020)
Emerg. vol. (other Sps)	0.047	-0.196^{***}	0.045^{*}	-0.123^{***}
	(0.026)	(0.026)	(0.020)	(0.019)
Cross-sectional effects				
Elect. vol. (focal Sp)	-0.049^{***}	0.034^{***}	-0.021^{***}	0.014^{**}
	(0.009)	(0.009)	(0.005)	(0.005)
Emerg. vol. (focal Sp)	-0.010	-0.144^{***}	0.008	-0.106^{***}
	(0.015)	(0.014)	(0.008)	(0.008)
Elect. vol. (other Sps)	$0.059^{\dagger}$	0.168^{***}	0.003	0.057^{**}
	(0.029)	(0.029)	(0.019)	(0.019)
Emerg. vol. (other Sps)	-0.037	-0.108^{**}	0.007	-0.023
	(0.032)	(0.032)	(0.021)	(0.021)
Observations	41,564		41,564	

$^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The results from this combined model specification are reported in Table EC.13, and are consistent with those reported in the main paper.

EC.8. Generating the dependent variable

Over the next two pages we provide an example demonstrating how the dependent variable is generated.

Generating the dependent variable

This example runs through how HRG-level costs are aggregated to the HRG-chapter level, using real data from the 2014-15 financial year. To demonstrate we will assume there are only two hospital trusts, RGT (Cambridge University Hospitals) and RTH (Oxford University Hospitals), and that HRG chapter of interest is "C" (corresponding to the Mouth, Head, Neck and Ears). To simplify further, within HRG chapter C we assume there are only two HRGs:

- CZ21V: Minor Head, Neck and Ear Disorders, with CC
- CZ21Y: Minor Head, Neck and Ear Disorders, without CC

The table below gives the volume, average cost, and average length of stay (in days) of patients in each trust, for each type of patient (DC = day case, EL = elective inpatient, EM = emergency inpatient), for each of the above HRGs.

Hospital	HRG	Patient type	Volume	Average cost	Average LoS
RGT	CZ21V	EL	7	1,788	2.28
RGT	CZ21V	EM	611	646	1.56
RGT	CZ21Y	DC	14	340	1
RGT	CZ21Y	EL	3	2,171	2
RGT	CZ21Y	EM	152	544	1.11
RTH	CZ21V	DC	6	928	1
RTH	CZ21V	EL	4	3,351	2
RTH	CZ21V	EM	645	623	1.31
RTH	CZ21Y	DC	3	871	1
RTH	CZ21Y	EL	8	2,857	1.5
RTH	CZ21Y	EM	209	559	1.11

We now describe how the dependent variable is constructed from the above.

Preparing the data

1. First, we combine DC and EL patients by taking their weighted-average – e.g., the average cost of DC-EL patients with HRG CZ21V in RTH is equal to (6*928+4*3,351)/10 = 1,897 – to form the new table below:

Hospital	HRG	Patient type	Volume	Average cost	Average LoS
RGT	CZ21V	DC-EL	7	1,788	2.28
RGT	CZ21V	EM	611	646	1.56
RGT	CZ21Y	DC-EL	17	663	1.18
RGT	CZ21Y	EM	152	544	1.11
RTH	CZ21V	DC-EL	10	1,897	1.40
RTH	CZ21V	EM	645	623	1.31
RTH	CZ21Y	DC-EL	11	2,315	1.36
RTH	CZ21Y	EM	209	559	1.11

2. We also combine the across-hospital data and calculate for each HRG–admission-type combination: (a) the percentage of patients of that admission-type allocated to that HRG, (b) the average cost of treating those patients, and (c) their average length of stay. This produces the table below.

HRG	Patient type	% of total volume	Combined average cost	Combined average LoS
CZ21V	DC-EL	37.8%	1,852	1.76
CZ21V	EM	77.7%	634	1.43
CZ21Y	DC-EL	62.2%	1,312	1.25
CZ21Y	EM	22.3%	553	1.11

Case-mix adjustment

3. To case-mix adjust, we first take the across-hospital % of total volume associated with each HRG and multiply this by the average cost/LoS in each hospital. For example, the case-mix adjusted cost of HRG CZ21V at RGT is equal to 1,788*0.378 = 675. This results in the following:

Hospital	HRG	Patient type	% of total volume	Case-mix adjusted cost	Case-mix adjusted LoS
RGT	CZ21V	DC-EL	37.8%	675	0.86
RGT	CZ21V	EM	77.7%	502	1.21
RGT	CZ21Y	DC-EL	62.2%	413	0.73
RGT	CZ21Y	EM	22.3%	121	0.25
RTH	CZ21V	DC-EL	37.8%	717	0.53
RTH	CZ21V	EM	77.7%	484	1.02
RTH	CZ21Y	DC-EL	62.2%	1,441	0.85
RTH	CZ21Y	EM	22.3%	125	0.25

Aggregating costs to the HRG chapter level

4. The next step is to take the sum of the case-mix adjusted costs in each hospital for each admission-type. This is equal to the chapter level average cost per patient (i.e., the cost of treating an `average' patient in that hospital). For example, the average cost of an `average' DC-EL patient at RGT is equal to 675+413 = 1,088.
5. We also calculate the `expected' average cost of treating an `average' patient. This equals the sum of the case-mix weighted "combined average costs" from the table in (2.), e.g. for DC-EL patients is equal to (0.378*1,852 + 0.622*1,312) = 1,516. Putting this and the output from (4.) into a table gives:

Hospital	HRG chapter	Patient type	% of total volume	Avg. cost - chapter level	Avg. LoS - chapter level	Exp. cost - chapter level	Exp. LoS - chapter level
RGT	C	DC-EL	100.0%	1,088	1.59	1,516	1.44
RGT	C	EM	100.0%	623	1.46	616	1.36
RTH	C	DC-EL	100.0%	2,157	1.38	1,516	1.44
RTH	C	EM	100.0%	609	1.27	616	1.36

6. Finally, we divide the chapter level total cost/LoS at each hospital through by the expected total cost/LoS to generate a case-mix adjusted cost and LoS index for each patient type. These indices are the dependent variables used in our analysis.

Hospital	HRG chapter	Patient type	% of total volume	Cost index	LoS index
RGT	C	DC-EL	100.0%	0.72	1.10
RGT	C	EM	100.0%	1.01	1.07
RTH	C	DC-EL	100.0%	1.42	0.95
RTH	C	EM	100.0%	0.99	0.93

Notes

- In Step 2, the % of patients in each HRG and the combined average cost/LoS is instead determined from a set of 116 reference trusts (rather than the set of all trusts). These reference trusts are the set of trusts that are present in our data in each of the 10 years. This ensures that the case-mix is relatively stable over time. The exception to this is when the focal trust is one of the reference trusts. In this case, the combined average cost/LoS for that focal trust is instead calculated over all reference trusts *except* for the focal trust. This ensures that the numerator (hospital specified average chapter level cost/LoS) and denominator (expected average chapter level cost/LoS over the set of reference trusts) are independent in Steps 5/6.
- When an HRG is not present in the numerator of the cost/LoS indices – which can occur if a patient with that HRG is not treated in that hospital in that year – then the chapter level avg. cost/LoS calculated in Step 4 will be lower than in other hospitals. We thus also need to deflate expected avg. cost/LoS. To achieve this, we simply do not include that HRG when summing to calculate expected cost/LoS in Step 5. We also keep track of the % of the `average patient' that is observed in each hospital (which in our example is 100% in all cases). This becomes another control in our analysis.

EC.9. Difference between longitudinal and cross-section effects

First, it is important to note that the two effects capture distinct phenomena. This point was made in recent paper published in SMJ titled "A tale of two effects: Using longitudinal data to compare within- and between-firm effects" (Certo et al. 2017). In the managerial summary of the paper, the authors write: "Strategy research examines two sources of variation over time: what is occurring within the firm (e.g., Do firms perform better over time when investing more in R&D?) and what is occurring between firms (e.g., Do firms investing more in R&D outperform firms investing less in R&D?). [...] Our article highlights the benefits of theorizing and testing these two sources of variance, providing scholars the ability to broaden both the theoretical and empirical contribution of their research. This distinction is important to how research informs managerial decision making." Translating the R&D examples above into our context gives the following two questions about the sources of variation in hospital costs:

1. Do hospital costs decrease over time as they increase the volume of patients that they treat?
2. Do hospitals that have a higher volume of patients operate at lower cost than hospitals that have a lower volume of patients?

In this paper we ask these questions at the level for each specialty and admission type, allowing volume to differ along four dimensions (same specialty and type, different specialty same type, same specialty different type, different specialties and type).

If we assume that assets are frozen over the observation period,[9] then the first question above become effectively: as hospitals increase the volume of patients that they treat using the same set of assets, do hospital costs reduce? It is clear that any impact of volume on cost in these circumstances would be predominantly a utilization effect: treating a higher volume of patients with the same assets would indicate that the hospital is utilizing those assets more effectively. On the other hand, if a hospital that has a higher volume of patients is able to operate at lower cost then this is an indication of scale/scope economies. It turns out that question 1. is measured using the longitudinal (within) volume measures, while question 2. is measured using the cross-sectional (between) volume measures. This is why the cross-sectional (between) volume measures are the focus of this study.

Example

To make the points made above more concrete, we provide an example below. Suppose we have only two hospitals, A and B, and one specialty, e.g. the nervous system, and that hospitals A and

[9] This is a reasonable assumption given the 2008 economic crisis. Nevertheless, in EC.6.5 we show that our results are not overly sensitive to this assumption.

B experience no changes in capacity over an e.g. 5 year observation period. Hospital A treats the same number of elective patients in each of the 5 years, say 100. Hospital B, meanwhile, treats 140 electives in year 1 which increases by 5 each year until in year 5 they are treating 160 elective patients. Suppose that the relative cost of elective care at hospital A remains the same in each year, taking value 1 (i.e. equal to the average), while the relative cost of elective care at hospital B decreases by 0.025 per year from 1.0 in year 1 to 0.9 in year 5. Are there economies of scale?

We can try answer the above one of two ways, either (i) by looking at volume differences across hospitals, or else (ii) volume changes within hospitals. Let's say that we use (ii). Then since only hospital B exhibits volume changes over time, we must rely on hospital B only to estimate the scale effects. The data above woulds suggest that every 5 unit increase in elective volume decreases relative costs by 0.025, i.e. possible evidence of economies of scale. But recall capacity is fixed, so this isn't really capturing benefits associated with scale. Instead this is measuring an improvement in capacity utilization over time, i.e. hospital B is able to make better use of its resources to treat more patients for the same amount of capacity, and so is also able to reduce per patient cost. However, observe that hospital B *is* larger overall than hospital A, over the 5 years the average volume at hospital A is 100 while at hospital B it is 150. How then can we determine whether the larger scale of hospital B translates into reduced costs above and beyond the utilization gains hospital B achieves?

In order to identify economies of scale we must instead compare volume across hospitals *while controlling for utilization changes* (i.e. variation in volume) within a hospital over time. To account for these utilization changes we can use the longitudinal volume measures discussed above. Specifically, the average volume at hospital A over the 5 year period is 100, and at hospital B it is 150. Taking the differences between the volume of patients in any year and the average gives the longitudinal volume measure. This is equal to 0 at hospital A in each year, since volume does not change over time. At hospital B this is equal to -10 in year 1, increasing to $+10$ in year 5. The relative cost and longitudinal and cross-sectional volume are given in Table EC.14 below.

Controlling for utilization changes is equivalent to comparing the two hospitals when the longitudinal volume measures are set equal in value. This occurs when longitudinal volume is equal to 0 in hospitals A and B, or when relative cost and A is equal to 1.0 and at B is equal to 0.95 as shown in Table EC.14. Thus, even after accounting for changes in utilization over time there is still a cost difference between hospitals A and B. But hospital B operates at a larger scale than hospital A, so perhaps some of the cost difference can be explained as a function of this. In fact, this is exactly the scale economies that we are trying to identify, i.e. how costs differ across

Table EC.14 Example demonstrating difference between longitudinal and cross-sectional volume.

	Year				
	1	2	3	4	5
Relative cost, hospital A	1.0	1.0	1.0	1.0	1.0
LT elective volume, hospital A	0	0	0	0	0
CS elective volume, hospital A	100	100	100	100	100
Relative cost, hospital B	1.0	0.975	0.95	0.925	0.9
LT elective volume, hospital B	-10	-5	0	5	10
CS elective volume, hospital B	150	150	150	150	150

LT corresponds to longitudinal volume, CS corresponds to cross-sectional volume.

hospitals that operate at different levels of volume! In this example, hospital B treats 50 elective patients per year more than hospital A on average. This suggests that an additional 50 patients can reduce cost from 1.0 to 0.95, i.e. each additional patient a hospital treats reduces relative cost by 0.001. Note that in reality we control through our data and panel structure for many other factors (both observed and unobserved) that may drive differences in costs, and also use over 150 hospitals rather than just 2 to estimate this relationship, as well as aggregating over 16 distinct specialties.

Identification of effects of interest

Note that the formulation described above does not 'control out' from the between-hospital (cross-sectional) volume measures any of the possible drivers of the scale effects. For example, the fact that hospital B has a higher volume than hospital A which may confer advantages associated with e.g. learning, utilization[10], etc., still exists. Thus our theory and the measure that we use to capture it are consistent.

EC.10. Literature Review

A contribution of our work is to explore the question of whether, from an efficiency standpoint, scheduled elective activity and unscheduled emergency activity should be coproduced within the same general hospital, and also whether there are productivity spillovers between different medical specialties (specialties). This is an important question as spillovers across these dimensions are highly relevant for the current debate on business model innovation in regional hospital systems.

We can find only one other paper that considers economies of scope between patients of different admission type (emergency versus elective), though this unpublished study concludes that "For the elective dimension, methodological problems may be large enough to cast doubt on the validity

[10] We control with our longitudinal volume measures for changes in utilization within a hospital over time, *not* differences in utilization across hospitals.

of the results" (Kittelsen and Magnussen 2003). The closest paper to ours, and also the most methodologically rigorous, is by Gaynor et al. (2015). This paper separates DRGs into primary, secondary and tertiary levels (approximately based on how widely they are provided, especially in teaching hospitals), and then examines whether there are economies of scope between medical specialties within each level, and whether there are economies of scope across levels. However, this study uses only a single year of data and 324 data points, taking hospital level total annual operating expenses to be the dependent variable. This study is therefore conducted at both a higher unit of analysis than ours and lacks a panel data structure, and therefore suffers from the weaknesses laid out in our paper.

Over the next few pages we provide a summary of the results of our literature search.

Overview of literature on economies of scope in hospitals

Study #	Approach	Dependent variable	# obs.	Data structure, within/between	Unit of analysis	Scale effect	Scope effect	Case-mix adj.
This study	Multilevel model, accounts for omitted variable bias	Annual relative cost index	21,037	Panel (10 years), decompose into both within- and between-effects	Admission type (emergency or elective) and medical service line (e.g. musculoskeletal) level	Volume effect from patients of same admission type and from same service line (SL)	Volume spillovers from patients of (1) Same type other SLs, (2) Other type, same SL, (3) Other type, other SLs	HRG/DRG level direct cost adjustment, plus hospital-SL level random effects
1	Regression	Total annual variable cost	597	Single year study, between effects	Hospital level	Volume effect of (i) total inpatient and (ii) outpatient activity	Interaction term between inpatient and outpatient activity	Adjusted length of stay
2	Stochastic frontier analysis	Total annual operating cost	<200 (unclear)	Panel (4 years), pooled study	Hospital level	Volume effect of (i) ambulatory, (ii) ED, and (iii) inpatient cases	Pairwise interactions between the three types of activity in the "Scale effect" column	Weighted average of US Patient Management Category scores
3	Seemingly unrelated regression	Total annual variable cost	201 pairs	Single year matched pairs of firms that merged and those that did not, between effects	Hospital level	Differential effects of (i) acute (ii) intensive care (iii) sub-acute and (iv) outpatient activity on total variable cost between merging hospitals and non-merging hospitals	Pairwise interactions between the four types of activity in the "Scale effect" column	None
4	Wilcoxon matched-pairs signed ranks test	Total annual hospital expenses per admission	28 triples	Single year matched triple of three provider types (i) short-acute svcs. only, (ii) psychiatric svcs. only, (iii) both	Hospital level	N/A	Test for whether combined psychiatric and short-acute services cheaper than separate	None
5	Regression	Total annual operating expenses of the hospital	296	Single year study, between effects	Hospital level	Volume effect of (i) medical-surgical, (ii) pediatric, (iii) obstetric, (iv) ER and outpatient, and (v) "other" discharges	Pairwise interactions between the five types of activity in the "Scale effect" column	HCFA Medicare case mix index
6	DEA	Various measures of total annual	50	Single year study, between effects	Hospital level	N/A	Pairwise interactions between (i) medicine, (ii) surgery, (iii) gynecology, (iv) pediatrics	None

		outputs / inputs						
7	DEA	As above	~70 diversified hospitals, 60-80 specialized hospitals	Two years, with frontiers estimated separately	Hospital level	Compare scale efficiency of diversified and specialized hospitals	Identify whether there exist diversification economies	None
8	Regression	Total annual operating expenses	421	Panel (2 years), pooled effects	Hospital level	Effects of (i) primary/secondary (ii) tertiary (iii) chronic (iv) ambulatory activity	Pairwise interactions between the four types of activity in the “Scale effect” column	Adjust each case of Resource Intensity Weighted (RIW) cases
9	Random effects model	Total annual DRG-derived production value	160	Panel (3 years), pooled effects	Production unit level (inpatient, outpatient, ER)	Number of beds	Cost savings associated with the joint production of inpatient, outpatient, and ER activities.	DRG-based case-mix adjustment
10	DEA	Various measures of total annual outputs / inputs	467	Panel (8 years), pooled effects	Hospital level	N/A	Efficiency advantages of being specialized in elective vs emergency, surgical vs medical, outpatient vs other	DRG-weighted visit numbers for outputs
11	Regression	Total annual operating expenses	867	Single year	Hospital level plus production unit level (inpatient, outpatient, ER)	Effects of (i) acute (ii) intensive care (iii) sub-acute (iv) outpatient and (v) ambulatory activity	Interaction between outpatient and inpatient activities, and ambulatory and inpatient activities	Control for % patients in various medical specialties
12	Seemingly unrelated regression	Total annual variable costs	534	Panel (3 years), pooled effects	Hospital level	Effect of (i) inpatient, (ii) outpatient, (iii) maternity, (iv) emergency, and (v) surgery volume	Effect of join production of the five effects in the “Scale effect” column	HCFA Medicare case mix index, % ICU patients, % Medicaid
13	Regression	Total annual operating expenses	138	Single year, between effects	Hospital level	Effect of (i) ER, (ii) medical-surgical inpatient, (iii) pediatric, (iv) maternity, (v) other volume	Pairwise interactions between the five types of activity in the “Scale effect” column	None
14	GEE estimation	Total annual costs	4,793	Panel (11 years), pooled effects	Hospital level	Number of inpatient discharges and outpatient visits	Compares cost of production in a single specialty hospital versus general hospital	Medicare inpatient case-mix index
15	Correlated random	Total annual facility	1,733	Panel (5 years), pooled effects	Hospital level	Number of inpatient discharges and outpatient visits	N/A	Medicare inpatient case-mix index and

	effects model	operating expenses						average length of stay
16	Regression	Total annual variable costs	76	Single year, between effects	Hospital level	Effect of (i) acute surgical/medical inpatients and (ii) “other” volume	Interaction between acute surgical/medical inpatients and “other” volume	None
17	Leontief input – output model with random effects	Total annual operating costs	540 (?)	Panel (6 years), pooled effects	Hospital level	Volume of inpatient and outpatient services	Estimate how costs would change if inpatient and outpatient services were produced separately	Medicare inpatient case-mix index
18	Seemingly unrelated regression	Total annual operating expenses	324	Single year, between effects	Hospital level	Uses an output-adjusted measure of patient quantity to estimate optimal hospital size	Estimate scope (i) across medical specialties within tertiary, secondary and primary care and (ii) between tertiary, secondary, primary and outpatient care.	Case mix adjust outputs using a large number of controls and DRG related info
19	DEA	Various measures of operational economic efficiency	435 (?)	Panel (5 years), pooled effects	Hospital level	Compares efficiency of hospitals based on number of beds	Construct specialization index based on the degree to which hospital focused on one of the following: general medicine, surgery, psychiatric, emergency departments, intensive, and coronary care units	Average length of stay
20	Regression	Average inpatient charges multiplied by cost-to-charge ratios	1,735	Single year, between effects	Hospital-comorbidity level, where there are 3 comorbidity levels	Total number of discharges	The proportion of total discharges from the largest major diagnostic category	Medicare inpatient case-mix index
21	Stochastic frontier analysis	Total annual hospital costs	1,018	Panel (7 years), pooled effects	Hospital level	Total number of discharges and number of outpatient visits	Interaction between number of discharges and number of outpatient visits	APR-DRG case-mix index

Study #	Reference
1	Weaver, M. & Deolalikar, A. (2004). Economies of scale and scope in Vietnamese hospitals. Social Science & Medicine, 59(1), 199-208.
2	Wagstaff, A. & Lopez, G. (1996). Hospital costs in Catalonia: a stochastic frontier analysis. Applied Economics Letters, 3(7), 471-474.
3	Sinay, U. A., & Campbell, C. R. (1995). Scope and scale economies in merging hospitals prior to merger. Journal of Economics and Finance, 19(2), 107-123.
4	Rozek, R. P. (1988). A nonparametric test for economies of scope. Applied Economics, 20(5), 653-663.
5	Vita, M. G. (1989). Exploring hospital production relationships with flexible functional forms. Journal of Health Economics, 9(1), 1-21.
6	Prior, D. (1996). Technical efficiency and scope economies in hospitals. Applied Economics, 28(10), 1295-1301.
7	Prior, D., & Solà, M. (2000). Technical efficiency and economies of diversification in health care. Health Care Management Science, 3(4), 299-307.
8	Preyra, C., & Pink, G. (2006). Scale and scope efficiencies through hospital consolidations. Journal of Health Economics, 25(6), 1049-1068.
9	Kristensen, T., Olsen, K. R., Kilsmark, J., Lauridsen, J. T., & Pedersen, K. M. (2012). Economies of scale and scope in the Danish hospital sector prior to radical restructuring plans. Health Policy, 106(2), 120-126.
10	Kittelsen, S., & Magnussen, J. (2003). Economies of scope in Norwegian hospital production: A DEA analysis. Health Economics Research Programme. Working Paper.
11	Grannemann, T. W., Brown, R. S., & Pauly, M. V. (1986). Estimating hospital costs: a multiple-output analysis. Journal of health economics, 5(2), 107-127.
12	Fournier, G. M., & Mitchell, J. M. (1992). Hospital costs and competition for services: a multiproduct analysis. The Review of Economics and Statistics, 74(4), pp. 627-634.
13	Cowing, T. G., & Holtmann, A. G. (1983). Multiproduct short-run hospital cost functions: empirical evidence and policy implications from cross-section data. Southern Economic Journal, 49(3), 637-653.
14	Carey, K., Burgess, J. F., & Young, G. J. (2015). Economies of scale and scope: the case of specialty hospitals. Contemporary Economic Policy, 33(1), 104-117.
15	Carey, K. (1997). A panel data design for estimation of hospital cost functions. The Review of Economics and Statistics, 79(3), 443-453.
16	Scott, A., & Parkin, D. (1995). Investigating hospital efficiency in the new NHS: the role of the translog cost function. Health Economics, 4(6), 467-478.
17	Li, T., & Rosenman, R. (2001). Estimating hospital costs with a generalized Leontief function. Health Economics, 10(6), 523-538.
18	Gaynor, M. S., Kleiner, S. A., & Vogt, W. B. (2015). Analysis of hospital production: an output index approach. Journal of Applied Econometric, 30(3), 398-421.
19	Fragkiadakis, G., Doumpos, M., Zopounidis, C., & Germain, C. (2014). Operational and economic efficiency analysis of public hospitals in Greece. Annals of Operations Research, 2(247), 787-806.
20	Clark, J. R. (2011). Comorbidity and the limitations of volume and focus as organizing principles. Medical Care Research and Review, 1077558711418520.
21	Carey, K., Burgess, J. F., & Young, G. J. (2008). Specialty and full-service hospitals: a comparative cost analysis. Health Services Research, 43(5 pt. 2), 1869-1887.

References

Baltagi B, Wu P (1999) Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* 15(06):814–823.

Bell A, Johnston R, Jones K (2014) Stylised fact or situated messiness? The diverse effects of increasing debt on national economic growth. *Journal of Economic Geography* 15(2):449–472.

Certo ST, Withers M, Semadeni M (2017) A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal* 38(7):1536–1556.

Gaynor M, Kleiner S, Vogt W (2015). Analysis of hospital production: an output index approach. *Journal of Applied Econometric* 30(3):398–421.

Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models* (New York, NY, USA: Cambridge University Press).

Kittelsen S, Magnussen, J (2003) Economies of scope in Norwegian hospital production: A DEA analysis. *Health Economics Research Programme* Working Paper.

NHS (2017) Overall patient experience scores. Technical report, Analytical Services, NHS England, https://www.england.nhs.uk/statistics/statistical-work-areas/pat-exp/, Accessed: 2017-07-17.