Who Should See the Patient? On Deviations from Preferred Patient-Provider Assignments in Hospitals

Mariam K. Atkinson[1], Soroush Saghafian[2]

[1]Department of Health Policy and Management, T.H. Chan School of Public Health, Harvard University, Boston, MA 02115
[2]Harvard Kennedy School, Harvard University, Cambridge, MA 02138

In various organizations including hospitals, individuals are not forced to follow specific assignments, and thus, deviations from preferred task assignments are common. This is due to the conventional wisdom that professionals should be given the flexibility to deviate from preferred assignments as needed. It is unclear, however, whether and when this conventional wisdom is true. We use evidence on the assignments of generalist and specialists to patients in our partner hospital (a children's hospital), and generate insights into whether and when hospital administrators should disallow such flexibility. We do so by identifying 73 top medical diagnoses and using detailed patient-level electronic medical record (EMR) data of more than 4,700 hospitalizations. In parallel, we conduct a survey of medical experts and utilize it to identify the preferred provider type that *should have been* assigned to each patient. Using these two sources of data, we examine the consequence of deviations from preferred provider assignments on three sets of performance measures: operational efficiency (measured by length of stay), quality of care (measured by 30-day readmissions and adverse events), and cost (measured by total charges). We find that deviating from preferred assignments is beneficial for task types (patients' diagnosis in our setting) that are either (a) well-defined (improving operational efficiency and costs), or (b) require high contact (improving costs and adverse events, though at the expense of lower operational efficiency). For other task types (e.g., highly complex or resource-intensive tasks), we observe that deviations are either detrimental or yield no tangible benefits, and thus, hospitals should try to eliminate them (e.g., by developing and enforcing assignment guidelines). To understand the causal mechanism behind our results, we make use of mediation analysis and find that utilizing advanced imaging (e.g., MRIs, CT scans, or nuclear radiology) plays an important role in how deviations impact performance outcomes. Our findings also provide evidence for a "no free lunch" theorem: while for some task types deviations are beneficial regarding some performance measures, they can simultaneously degrade performance in terms of other dimensions. To provide clear recommendations for hospital administrators, we also consider counterfactual scenarios corresponding to imposing the preferred assignments fully or partially, and perform cost-effectiveness analyses. Our results indicate that enforcing the preferred assignments either for all tasks or only for resource-intensive tasks is cost-effective, with the latter being the superior policy. Finally, by comparing deviations during weekdays and weekends, early shifts and late shifts, and high congestion and low congestion periods, our results shed light on some environmental conditions under which deviations occur more in practice.

Key words: Hospital Operations; Task Assignment; Patient-Provider Assignment; Division of Labor
History: Version: November 21, 2022.

1. Introduction

Organizations often implement coordination mechanisms to formalize processes and move toward more preferred practices, which can aid in reducing unnecessary variability and improve performance. In the professional setting, professionals have commonly held views about how coordination mechanisms should manifest in their daily practice. Through such coordination mechanisms, the role of preferred practices versus those which deviate can have important bearing on performance outcomes. For example, some professionals tend to deviate from preferred practices in batching and sequencing tasks for completion, which ultimately diminishes productivity (Ibanez et al. 2017). In other examples, the loss of preferred practices due to the transfer of a central manager to a competing organization diminishes performance (Briscoe and Rogan 2015, Aime et al. 2010), and deviations from preferred assignment of medical/surgical patients to hospital units worsens the outcomes (Song et al. 2020).

Realizing the tension between enforcing preferred practices and allowing flexibility among professionals to deviate when needed, some theories suggest an "optimal" level of upholding preferred practices (Engel 1969). In contrast, others argue that such inflexibility can be disastrous, since professionals must regularly deal with complex and nuanced issues in serving their clients (Champy 2009). Similarly, while decision theories discuss that decision-makers must possess the ability to know when to deviate in their practice (see, e.g., D'Adderio 2014, Cyert and March 1963, and the references therein), operations management theories suggest avoiding deviations, indicating that standardizing how tasks are routed to different servers have important advantages (for studies on optimal routing in service systems with heterogenous servers, see, e.g., Armony and Ward 2010, Saghafian et al. 2022a, and the references therein). Some studies also provide clear evidence that standardizing how tasks (e.g., arriving patients) are routed to servers (e.g., providers) through implementing specific patient-provider assignment algorithms can bring various benefits to hospital operations (see, e.g., Traub et al. 2016). Acknowledging these conflicting arguments, it is unclear whether and when it is beneficial to enforce preferred task assignments and thereby remove any potential deviation. More broadly, as the healthcare sector is moving towards increased levels of transparency through efforts such as publicly reporting hospitals' outcomes (see, e.g., Saghafian and Hopp 2019, 2020), it is becoming more important to understand whether by eliminating workarounds (see, e.g., Spear and Schmidhofer 2005, Tucker et al. 2020) hospitals can improve their performance measures. Improved level of performance, in turn, might reduce some undesired but common and large-scale events in the healthcare sector, including hospital closures (see, e.g., Saghafian et al. 2022c) and unnecessary vertical integration (see, e.g., Saghafian et al. 2022b).

We are particularly motivated by the assignment of generalists and specialists in our study hospital—a children's hospital on the west coast of the United States. The hospital's administrators were interested to know whether they should (a) develop guidelines on the assignment of generalist versus specialist physicians (for each arriving patient with certain medical conditions) based on the preferred practice as represented by consensus opinion of medical experts, and (b) enforce such assignment guidelines in their hospital to remove any potential deviation. Our study allows our partner hospital, in addition to many other hospitals dealing with similar issues, to gain a deep understanding of the impact of developing and enforcing preferred assignments (points (a) and (b) above) on various performance outcomes. Specifically, the central question in our study is: *What is the impact of following preferred assignments (identified based on the consensus opinion of medical experts) on various performance outcomes?* Conversely, *when is deviation from such preferred assignments beneficial?*

To answer these questions, we collected data representing nearly six years of electronic medical record (EMR) information from our partner hospital. Our data contained information on 4,729 hospitalized patients with common pediatric diagnoses for which physicians desired clearer guidelines around generalist and specialist assignment. Separately, in close collaboration with the hospital's administrators, we also conducted surveys and collected data by asking medical experts, including various physicians in our partner hospital, who *should have been* in charge of each patient given his/her documented diagnoses. We then made use of the consensus opinion of the surveyed experts to determine the *preferred assignment*. Thus, the term "preferred assignment" in our study refers to the most commonly-recommended opinion among medical experts regarding whether a specialist or a generalist should have been in charge. Similarly, we use the term "deviation" to refer to situations when the actual assignment differs from the preferred one. We adopt these definitions based on the goals of the administrators at our partner hospital who want to know whether developing and enforcing guidelines using the consensus opinion of medical experts, and thereby, removing any deviations from them will yield improvements.[1]

We analyzed the consequences of deviations from preferred assignment along three dimensions of performance: *operational efficiency* (measured by length of stay[2]), *quality* (measured by the occurrence of 30-day readmissions and adverse events), and *cost* (measured by total charges). We also assessed outcomes based on the interaction between the preferred assignment and task (i.e., diagnosis) characteristics, focusing on the following types of tasks: *well-defined*, *high complexity*, *high contact*, and *resource-intensive*. Finally, we (a) performed mediation analysis to understand the mechanism behind some of our results, (b) employed cost-effectiveness analysis on some implementable counterfactual policies to provide clear recommendations for hospital administrators on when and how to enforce guidelines to remove deviations from the preferred assignment, and (c) conducted various robustness checks (including running sensitivity analyses on how the consensus opinion is measured, making use of an instrumental variable (IV) approach, and applying the 1-nearest neighbor propensity score matching method) to further examine the soundness of our main findings.

Our analysis of the data suggests that, for certain types of tasks, deviations from preferred assignments are beneficial. Specifically, we find that for well-defined tasks such deviations can reduce costs and improve operational efficiency. For resource-intensive tasks, however, such deviations are

[1] It should be noted that when a deviation occurs in practice, it might be because of a judgement call made by the decision-maker. This can occur due to various reasons (e.g., personal interests, imposing professional power, or belief about capacity), studying which is beyond the scope of our work. Importantly, however, we note that deviations, regardless of the underlying reason behind the judgement call, impact performance outcomes. Thus, we aim to help hospital administrators by studying whether or not they should impose guidelines that can eliminate all such deviations.

[2] All else equal, a shorter average patient length of stay translates to a better throughput. As such, length of stay is a widely used measure for gauging operational efficiency of hospitals.

associated with higher costs. Further, when tasks involve high contact, deviations from preferred assignments are associated with worse operational efficiency, lower costs, and higher quality of care through a lower occurrence of adverse events. Put together, our findings indicate that deviating from preferred assignments is beneficial for some (but not all) tasks. In particular, we find that it is beneficial to deviate when (a) the patient's needs are well-defined (improving operational efficiency and cost), or (b) serving the patient requires high contact (improving costs and adverse events, though at the expense of lower operational efficiency). For other task types, we find that hospital administrators should develop and formally enforce guidelines to remove deviations from preferred assignments: such deviations are either detrimental or come with no tangible benefits.

To shed light on a potential causal mechanism behind our findings, we make use of mediation analyses. Our results show that use of advanced imaging (e.g., MRIs, CT scans, or nuclear radiology) provides an important causal channel through which deviations impact performance measures (operational efficiency, costs, and quality). This is partially because use of advanced imaging significantly differs between specialist and generalists, and whether advanced imaging is used or not can influence outcomes such as operational efficiency, costs, and quality. This mediating impact of advanced imaging, in turn, depends on patients' needs as identified by their task type.

Finally, we make use of our findings to provide clear recommendations for hospital administrators and assist them in making better decisions with respect to the underlying tradeoffs in enforcing preferred assignments. We do so by considering counterfactual scenarios corresponding to imposing the preferred assignments fully or partially, and by performing cost-effectiveness analyses. Our results indicate that enforcing preferred assignments either for all tasks or only for resource-intensive tasks is cost-effective, and that the latter—enforcing preferred assignments only for resource-intensive tasks—is the superior policy.

1.1. Contributions and Implications

Our research provides evidence on how hospitals can utilize a *task-type view* to standardize assignment decisions by understanding the impact of following the preferred assignment as detrained by medical experts. In doing so, our study offers new insights that have both managerial and theoretical implications, which we discuss next.

Using data from our partner hospital, we provide evidence of a significant amount of deviations in assigning generalists versus specialists to patients. From a managerial and practical perspective, this removes the perception that most patients are admitted by a specific type of physician (e.g., a specialist) and for a specific reason (e.g., surgery), and hence, there will be no disagreement about the type of the physician that should be in charge.[3] Furthermore, since such deviations affect various performance metrics such as quality, operational efficiency, and costs, a natural question

[3] For example, we observe that only for 5.3% of hospitalizations in our sample there is complete agreement among medical experts about the type of the physician (generalist versus specialist) that should have been in charge.

for hospital administrators is whether they should develop and enforce formal guidelines to remove deviations. We answer this question by showing that enforcing guidelines around preferred assignments can have advantages and disadvantages. Understanding the underlying tradeoffs for each task type—and knowing for which type to formally enforce such assignments as we show in this study—affords managers the opportunity to improve operational efficiency, cost, and quality of their services. Additionally, as our interviews at our field site reveal, creating formal assignment guidelines with the goal of improving outcomes in mind, and understanding when physicians should have the flexibility to deviate from such guidelines can be exceedingly helpful in practice:

> *"It would be helpful to first understand what the current process for assignment is and what the goals for assignment are. [...] My current dissatisfaction arises from not knowing what the goals and expectations are and thus not being able to adapt. These also do not seem to be consistent among all the hospitalist faculty that I rotate with."* [Gastroenterologist]

and

> *"If there are pre-determined guidelines in terms of how physician assignments are done, this would be a benefit to how clinical care would work efficiently [. . .] The 'guidelines' are a difficult task to accomplish, as it would be a hard to achieve a sort of universal consensus. Perhaps the best method is to have guidelines be guidelines, but to have open communication if there are questions or concerns regarding physician assignment."* [Endocrinologist]

Additionally, our results help hospital administrators as well as physicians gain a better understanding of the environmental conditions under which deviations occur more. In particular, our findings indicate that deviations are more frequent during weekends than weekdays and during morning shifts (8am-1pm) compared to other shifts. However, we observe that deviations occur similarly during high congestion (busy) and low congestion (less busy) periods.

Our results also have a few essential theoretical implications. First, to the best of our knowledge, our work is the first to shed light on the dependency between task type and whether allowing flexibility for deviations from preferred assignments can be beneficial. Second, by taking into account various performance metrics (operational efficiency, quality, and cost), we provide evidence for a "no free lunch" theorem: although for some task types (e.g., those requiring high contact) deviations are beneficial in certain aspects of performance (e.g., quality and cost), they are simultaneously detrimental in other aspects (e.g., operational efficiency). Thus, while enforcing preferred assignments is the dominant strategy for some tasks types, permitting deviations is typically not dominantly the better option, regardless of the task type. Third, while our study is focused on understanding the impact of deviations on hospitals' performance (and not the reasons or individuals behind deviations), our results point to future work that can further investigate the potential interplay between professional status and deviations from preferred assignments. For example, our data shows that specialists deviate from preferred assignments more than generalists. A reason for this could lie in the way professional status manifests in the workplace. Our aim with this research is to pave the way for future studies to investigate this and other potential reasons for deviations in professional practice.

2. Theoretical Background

2.1. Task Assignment between Generalists and Specialists

The known skillsets generalists and specialists provide in an organization can facilitate task assignment, or the mapping of tasks to different types of professionals (Puranam et al. 2014). For example, generalists provide a more holistic perspective with their breadth of knowledge while specialists provide more tailored services with their intricate and detailed knowledge (Cohen 2013, Currie and White 2012, Grant 1996). Consequently, whereas specialists' narrower expertise enables them to deliver more customized services to increase effectiveness in meeting particular client demands, generalists' broader expertise can facilitate the efficient completion of preferred client demands (Chase and Tansik 1983).

Although the operations management literature has studied optimal task assignment in settings where professionals differ in their knowledge levels and other abilities (see, e.g., Saghafian et al. 2018, for optimal task assignment rules in knowledge-based service systems), the above-mentioned differences as well as the significant overlap between generalist and specialist expertise bring new challenges to understanding suitable ways of task assignment. For example, generalists' and specialists' overlapping jurisdictions can make the process of task assignment considerably complex, since professional contexts increasingly embody collaborative environments in which gray areas around roles can become more salient as jurisdictions have greater opportunity to collide (Thornton et al. 2005).

In our study's setting, task assignment occurs between generalists and specialists who are at the organization-client interface and equipped with expertise to (a) make autonomous decisions in managing the complexity inherent in their daily work (Thomas and Hewitt 2011), and (b) proactively deviate from preferred practices to enhance service and accommodate circumstances in the work environment (Reay et al. 2006). Due to the sometimes unnecessary complications of task assignment, and the subsequent need to streamline professional work, organizations have attempted to developed and follow formal task assignment rules. Yet, professionals may still maintain their capacity for discretionary decision-making in their work.[4] Since professionals engage in preferred and deviating practices to manage task assignment, we aim to study whether and when deviation from preferred assignments can improve performance. In measuring performance, we take into account the fact that hospitals care about various metrics in a simultaneous way (see, e.g., Roth et al. 2019). Hence, we consider a holistic approach and study the impact of deviations

[4] Decision-making in clinical settings when there is ambiguity is often perplexing (see, e.g., Boloori et al. 2020), and thus, this capacity for discretionary decision-making may cause workarrounds and errors (see, e.g., Spear and Schmidhofer 2005).

from preferred task assignments on various dimensions of performance, including quality of care, operational efficiency, and costs.

Finally, we note that while little is known about when deviations from preferred task assignments can yield performance improvements, deviations have been studied in other settings, including task sequencing (see, e.g., Ibanez et al. 2017), task processing time (see, e.g., Hopp et al. 2007, Schultz et al. 1998, 1999), following system-generated recommendations (see, e.g., van Donselaar et al. 2010), and assigning medical/surgical patients to hospital units (Song et al. 2020). Similarly, previous studies have discussed the effect of differentiating between task types on improving performance, including separating complex patients (see, e.g., Saghafian et al. 2014) and customer types that should be routed to a specialist (see, e.g., Shumsky and Pinker 2003) or to a telemedical physician (see, e.g., Saghafian et al. 2018). Our research unifies many of these separately studied, yet critical aspects of professional performance.

2.2. Task Types and Hypotheses

Seminal literature on task design representing the works of classical theorists articulates four prominent task features, which are the focus of our research: *well-defined*, *high complexity*, *high contact*, and *resource-intensive*.

In what follows, we describe each of these task types that we considered in our setting and provide related hypotheses. We developed definitions of each of these task types such that they had relevance to our specific hospital context and simultaneously represented the theoretical basis of each task as described in the extant literature. These definitions were co-developed with four hospitalist physicians in our partner hospital, and were subsequently corroborated with two specialists for consistency. Similarly, we developed a separate hypothesis for each task category in collaboration with the administrators in our partner hospital and by considering their need to better understand the types of patients for which they should enforce preferred assignments.

Well-Defined Tasks. Tasks can be organized in terms of how well-defined they are, specifically in terms of two related components: epistemological clarity and invariability in procedures. Epistemology refers to the means for knowing the nature of something—what an entity is and how it came into existence. It explains how "cognitive subjects come to know the truth about a given phenomenon in reality" (Bodenreider et al. 2004). Since epistemology explains how knowledge can be incorporated into practice, it can be a term used to describe the scope of knowledge pertaining to an entity. In the work environment, a task with high epistemological clarity means that both the type of problem and its source can be understood and measured. Relatedly, the second feature of well-defined tasks is that procedures in handling the task are largely invariable. This means that knowledge is applied to tasks through procedures that have been tried and tested, yielding greater certainty in the content and the context of application. Since well-defined tasks are those which have high epistemological clarity and low variability in procedures, we hypothesize that they are relatively conducive to preferred task assignment, meaning that following the preferred assignment (instead of deviating from it) that is identified based on the consensus opinion of experts is likely to be advantageous to performance for well-defined tasks. Thus, our first hypothesis is:

HYPOTHESIS 1. *For well-defined tasks, following the preferred assignment (identified based on the consensus opinion of experts) improves performance outcomes.*

High Complexity Tasks. Classical organizational design theories identify two key features of the scope of complex work: variety and interdependence (Langfred and Moye 2004). Task variety refers to the number of exceptions, or different types of situations and problems, encountered while performing a task (Perrow 1967). Tasks with high variety have many exceptions, so cannot be easily standardized. One result of an increasing number of exceptions may be the need to invoke different types of expert knowledge and skills to creatively handle a novel situation. Since settings with high task variety require more flexibility, bureaucratic and rule-based structures are not as effective. The other feature of complex work is interdependence, which refers to tasks that rely on what others do. For such tasks, therefore, designating a professional from the mix can be a challenging endeavor (Thompson 1967). The highly complex task may either warrant the expertise of a specialist on a particular or rare subject matter, or the generalist if it requires a holistic view to be able to coordinate knowledge across multiple professional disciplines. Furthermore, these features of complex tasks can vary on a case-by-case basis, making any standardized assignment less effective. Thus, we hypothesize that these features of highly complex tasks would make them relatively not conducive to, and suitable for, enforcing preferred assignments. Specifically, if in deciding the appropriate provider type that should be assigned, professionals are forced to follow what the preferred assignment suggests, their assignment decisions might worsen, since they may miss the opportunity to account for the nuances of complex tasks. This leads us the following hypothesis:

HYPOTHESIS 2. *For high complexity tasks, following the preferred assignment (identified based on the consensus opinion of experts) worsens performance outcomes.*

High Contact Tasks. The extent of client contact, or the degree to which a client is in direct contact with a particular service facility relative to the total time needed to service the customer, is known to be an important factor that affects organizational performance (see, e.g., Chase and Tansik 1983). The presence of customers with high contact can disrupt the flow of work, and also put exaggerated demands on professionals that would not otherwise occur (Danet 1981). In contrast, when work involves serving customers requiring low contact, the service process involves less dependence on the conditions of the organizational environment (e.g., workload, staff available, overnight shift requirements) compared to high client contact settings (Chase and Tansik 1983). Therefore, assignment decisions are typically programmable for work involving low client contact.

For high client contact tasks, however, it is likely that programming assignments by defining predetermined rules such as those defined as preferred, and enforcing them through disallowing potential deviations can have negative consequences. Thus, we hypothesize that:

HYPOTHESIS 3. *For high contact tasks, following the preferred assignment (identified based on the consensus opinion of experts) worsens performance outcomes.*

Resource-Intensive Tasks. Resources refer to the inputs required to effectively complete a task, which for the purposes of this study, may include both human and physical capital. In the professional work environment, decisions about task partitioning involve efficient utilization of specialized resources (von Hippel 1990). For optimal performance, tasks that are highly resource-dependent require a systematic identification, selection, and assignment of resources (Crowston 1991, 1997). Thus, the degree of resource intensity can be an important determination of the type of professional who should be assigned to the task. In particular, when the level of resource intensity is high, it is often clear in practice who should be in charge of the task. We hypothesize that this makes such tasks suitable for preferred assignment, since for a given task a particular professional may have the best expertise in making assessments of resource requirements (i.e., resource variety that should be coordinated by a generalist or resource specificity that requires the expertise of a specialist) and subsequently managing the resources for completing the task at hand:

HYPOTHESIS 4. *For resource-intensive tasks, following the preferred assignment (identified based on the consensus opinion of experts) improves performance outcomes.*

2.3. A Potential Mechanism

Our research also explores a possible causal mechanism that could explain the relationship between preferred (versus deviating) assignment and performance outcomes. Specifically, we identify a potential mediating effect that could explain (a) why performance outcomes may be impacted by deviations from preferred assignments, and (b) why this impact depends on the task type. To identify the underlying potential mediator, we first explore the potential reasons why generalists and specialists might differ in their practices. In taking this approach, we make use of the substantial literature on how generalist and specialist physicians vary in utilizing resources (see, e.g., Auerbach et al. 2000, Boom et al. 2012, Greenfield et al. 1992, Harrold and Gurwitz 1999, Stevens et al. 2017). Specifically, we focus on the fact that generalist and specialist physicians differ in their use of advanced imaging (e.g., MRIs, CT scans, or nuclear radiology), and find that use of advanced imaging creates a mediating channel through which deviations impact performance measures. The strength of this mediation depends both on the task type and the performance outcome, which further explains the reason behind some of our main findings (see Section 4.2 for more details).

3. Research Setting, Data, and Analysis

3.1. Research Setting

We use data that we collected from an urban academic children's hospital on the west coast of the United States, and focuses on preferred versus deviating generalist and specialist physician assignments to patients. We focus on generalists and specialists mainly because of the needs of our partner hospital, and the fact that task assignments between generalists and specialists are often not clear-cut. Indeed, it is widely-known that the boundaries between the expertise of generalists and specialists is often blurred in practice, and hence, there is a considerable level of discretion on how tasks are assigned to them (Sinha and Van de Ven 2005). This gives us enough data points in which deviations have occurred, and in turn, allows us to study when enforcing preferred assignments and removing such deviations is beneficial. In our setting, the generalist physicians are hospitalists who work primarily in the inpatient environment. The specialist physicians included in our analysis belong to one of seven different specialties: cardiology, endocrinology, gastroenterology, hematology/oncology, neurology, pulmonology, and rheumatology.

3.2. Data

Using a mixed methods approach, we collected data from two primary sources: survey of physicians and EMR data.

Survey of Physicians. To identify preferred physician assignment, we administered an online survey using Qualtrics software to the department of pediatrics at the children's hospital, which included hospitalists and specialists belonging to the seven specialties defined above. To design the survey, we iteratively solicited feedback from the division head of hospital medicine and the research director at the hospital. Once a pilot version of the survey was developed, we made modifications based on the feedback we received after performing cognitive tests for question clarity and relevance on three hospitalists. The survey listed the "top diagnoses" for each specialty (a total of 176 diagnoses across seven specialty areas; see Table EC.9 in the Appendix for more details) and asked respective specialists, as well as a randomized group of hospitalists, who should be in charge: a generalist or a specialist.

The 176 medical diagnoses across the seven specialty areas listed in the survey did not include extraneous information elaborating on the context of care or other patient characteristics. Rather, we developed our survey questions much like how the standards of care are developed in guiding medical practice for certain diagnoses. Specifically, standards of care are typically informal or formal guidelines developed by specialty societies or organizations (e.g., Institute of Medicine, American College of Physicians) representing the majority expert opinion on the diagnostic, treatment, care process, and clinical practice pathway for patients with particular conditions (e.g., diabetes). The specific medical diagnosis serves as the basis for developing such rules and protocols around patient care by majority expert opinion (Jue et al. 2019, Qaseem et al. 2019). Avoiding complex contextual information besides the medical diagnosis in developing such rules enables hospitals to have guidelines that can be easily implemented in their practice. Thus, we also avoided detailed contextual information in our survey, and focused on medical diagnosis as its basis.

Our survey had a 44% response rate ($n = 66$ physicians which included 46 specialists and 20 hospitalists) and a 100% response rate across the eight divisions (seven specialties plus hospital medicine) surveyed. The 44% response rate ($n = 66$ physicians) was determined sufficient for two main reasons. First, based on Dillman's classical work on survey methods, and accepting a ± 10% sampling error, a minimum of about 55 responses in our setting would be sufficient to provide accurate results (see, e.g., Table 5.1. of Dillman (2011)), which is below our total number of responses ($n = 66$). Second, our analyses of respondents' characteristics, including position (faculty, fellow, contract), years of experience, gender, age, and ethnicity, among others revealed that the respondents' characteristics are not different than the general population of physicians at our partner hospital (see also Krikorian et al. (2018)). This gave us further confidence that our survey is unlikely to be subject to selection bias.

If the responding physician thought they would not be able to specify a generalist or specialist, the option "unsure—my selection depends greatly on other factors" was provided; that option was selected approximately 5% of the time and was dropped from our analysis. Top diagnoses lists for each specialty were first generated based on a query to the health information management department at the hospital, requesting the highest volume conditions for which patients were hospitalized. We also included variables for task category, or diagnosis type, which correspond to the task dimensions described earlier: well-known, high complexity, high contact, and resource-intensive. To operationalize the task types as a representation of diagnosis categories, we asked eight focus groups of two to three hospitalists each to classify the top diagnoses from the survey into four corresponding categories. The category definitions were inductively derived during an observational period at the hospital (e.g., patient rounds, physician meetings), consultation with the medical and management literature, and with input from several physicians at the hospital in order to ensure relevance to the professionals in our context.

Electronic Medical Records (EMRs). We also collected EMR data for patients hospitalized between January 1, 2009 and August 31, 2015 for any of 73 top diagnoses ($n = 4,729$ hospitalizations).[5] Since we chose common pediatric diagnoses for which physicians in our partner hospital desired more clarity around assignment, our sample represents a subset of all hospitalizations during the data collection period. The top diagnoses list aligning with the EMR data was shorter for multiple reasons, including issues with mapping distinct ICD-9 codes to the condition specified, as well as the fact that a modal response did not exist for several conditions. The EMR data included

[5] The number of diagnoses in our sample was less than that which was in our survey (176) because we excluded any diagnosis that occurred in fewer than 20 hospitalizations.

detail on patient demographics, the nature of the diagnoses, patient outcome measures, and the physician of record (i.e., the physician ultimately taking responsibility for the patient's care). Identification of documented adverse events was a more complicated process. Using the assistance of an EMR coder at the hospital, we examined the five primary ICD-9 codes associated with each patient hospitalization for evidence of any adverse events, and categorized them accordingly.

3.3. Analysis

We analyzed outcomes based on four performance metrics: *length of stay, total charges, 30-day readmissions, and adverse events*. For length of stay and total charges, we used a generalized linear model (GLM) with gamma distribution to account for non-negativity and skewedness of distributions of these continuous outcome variables. For readmission and adverse events, we used a GLM with binomial distribution, since these outcome measures are integers. For all GLM models, we assumed proportional effects for each outcome variable, and hence, applied a logarithmic link function. We addressed the clustering of patients by physician using robust standard errors, and performed cluster correction based on the physician assigned. Further, about 7.5% ($n = 356$) of the hospitalizations in our sample had patient-physician assignments that repeated more than once (i.e., repeat patients assigned to the same physician across their different hospitalizations). After checking for differences in our results if we excluded these cases, we found that the impact of excluding these patients is nominal; thus, we kept these hospitalizations in our analysis. Our main effects of interest were preferred (versus deviating) assignment and type of task (well-defined, high complexity, high contact, and resource-intensive). Our control variables included various patient, physician, diagnosis, and hospital condition characteristics as well as year fixed effects. Below, we discuss all of these in detail.

Dependent Variables. Our outcomes variables focused on three dimensions of performance: *operational efficiency*, *cost*, and *quality*. Operational efficiency was represented by *length of stay*, which is the total number of days from when a patient is admitted into the hospital until s/he is discharged (see also footnote 1). To measure cost, we used *total charges*, which is the amount billed to insurance for costs incurred during the patient's hospitalization. Of note, total charges are an estimation of costs, since this amount is billed but may not be the amount ultimately reimbursed and/or incurred. Finally, we measured quality by making use of two metrics: (1) *30-day readmission*, which captures the number of patients that were hospitalized again within thirty days of their last discharge (a binary variable), and (2) *adverse events*, which represents the number of patients that had a non-surgical harmful event resulting from care at the hospital (a binary variable). Adverse events were identified initially by the physician in charge of the patient, and then were revised as needed after medical record coders at the hospital conduct a review of the patient's hospitalization records post-discharge. These include adverse drug events, infections, and device events.

Independent Variables. Our independent variables are described below. Note that, as we described earlier, our use and definition of these variables (e.g., task types) are based on (a) our collaboration with the expert physicians in our partner hospital, and (b) available studies in the literature.

1. *Preferred assignment* is a binary variable that equals one if the physician assigned (a generalist or specialist, as recorded in the medical records) matches who should have been assigned based on the modal results from the physician assignment survey. In other words, preferred physician assignment reflects the survey responses, where the majority of physicians indicated that either a generalist or specialist should be assigned to a patient with a given condition.[6] For the 73 diagnoses in the survey, the responses indicated that specialists should be assigned to 44 (60%) diagnoses and generalists to 29 (40%) diagnoses.

2. *Well-defined diagnosis* was categorized according to the following definition: "well-defined expected course, complications, treatment and monitoring needs that are in a certain [physician's] domain of knowledge, skills and comfort-level." Such a diagnosis is less ambiguous, and is usually associated with what is known as the "standard of care" in the medical field, which is comprised of treatment guidelines that specifies appropriate patient care based on scientific evidence or collaboration among relevant medical professionals. The standard of care outlines patient treatment for a particular condition, such that medical errors and possible malpractice issues could be avoided. Thus, these diagnoses have a clearly outlined course of treatment and often coincide with legal protections of physician practice.

3. *High complexity diagnosis* was defined as follows: "patient has a diverse set of conditions and multisystem disease; may be technology dependent; has frequent inpatient admissions; and requires multiple medications, multiple specialists, and optimal care coordination across inpatient/outpatient settings" (Simon et al. 2010, Feudtner et al. 2014). Such a diagnosis involves multiple organ systems as well as the ability to address uncertainties in the patient's diagnosis and course of treatment that result from higher levels of complexity in the underlying condition. High complexity conditions can be in the domain of either a specialist or generalist. A hospitalist may be appropriate for this type of condition to manage and coordinate multiple knowledge bases using a more holistic approach, or a specialist may be appropriate if such a condition is due to a particular underlying organ/system issue that requires their depth of expertise.

4. *High contact diagnosis* was defined as follows: "patient has a condition that requires frequent intervention and has a propensity for acute deterioration, and who is likely to require a physician who can be rapidly available." Such diagnoses require greater patient contact for multiple reasons.

[6] In our robustness checks, we alter this majority rule approach for measuring the consensus opinion, and make use of different definitions for identifying preferred assignment, including utilizing Fleiss' kappa tests as well as varying the 50% threshold used to define majority (see Section 5.3).

This type of patient is in an unstable state in which unexpected deterioration can rapidly take place, and therefore instinctive decision-making under conditions of uncertainty places greater demands on the physician in charge. Also, a patient with this type of diagnosis requires much time from their assigned physician, who may frequently intervene during the course of treatment.

5. *Resource-intensive diagnosis* was classified based on the following definition: "diagnosis/workup often requires use of multiple ancillary services and support (e.g., physical/occupational/speech therapy, social work, discharge planning, etc.), possible frequent admissions, and longer length of stay." The use of multiple types of resources, as well as the higher costs incurred from potentially longer patient stays and frequent readmissions, makes these types of diagnoses more resource-intensive.

Interactions. Since it is likely that the effect of following the preferred assignment on performance depends on the type of task category, we included interactions between each of the four task categories and the preferred assignment.

Controls. We controlled for several patient characteristics, including variables related to demographic and diagnosis characteristics. With regard to patient demographic characteristics, we included (1) age, defined as the patient's age at the time of discharge from the hospital; (2) sex, defined as the patient's sex (male or female) based on the EMR; and (3) insurance type, defined as a binary variable indicating whether the patient possesses private or public insurance. To control for the nature of the patient's condition, we incorporated the *chronic condition indicator (CCI)*, which is a case mix adjustment categorical variable taking a value 0 through 4 that dichotomizes ICD-9 codes into chronic or non-chronic conditions and aggregates chronic conditions into 1 of 18 mutually exclusive clinical groups to assess both the severity and complexity (i.e., number of comorbidities, or different diagnoses afflicting the patient) associated with the patient hospitalization (Khan et al. 2015, Berry et al. 2013). The CCI measure is an acceptable measure for characterizing a patient's condition, and can be used in place of diagnosis or DRG codes (see, e.g., Atkinson et al. 2018). Additionally, we controlled for timing of the patient's hospitalization, namely if it occurred during flu season, by using a binary variable indicating if the hospitalization occurred between October and April. During these months, hospitals typically experience higher patient volumes and patients who are sicker because their underlying conditions can be complicated by flu viruses or other seasonal illnesses. Patients hospitalized during this time are affected by more limited hospital resources and exposure to more sick patients. We also included volume during the hospitalization (i.e., the number of other patients hospitalized for a similar condition during each hospitalization) as one of our controls. In addition, we used year fixed effects, which includes controls for the year of the patient's hospitalization. This was done for multiple reasons, including the fact that an increasing number of hospitalists were hired at the hospital since 2009. Moreover, we also controlled for the time-of-day in our analyses. Finally, to control for the type of professional, we included a variable termed type of physician assigned (specialty), which reflects each of the different specialties in our analysis (e.g., hospital medicine, cardiology, endocrinology, etc.).

3.4. Potential Endogeneity, Hidden Confounders, and Other Limitations

We have made use of detailed patient level data to control for various factors that can affect our results. This has made us relatively confident that observable variables are not making our estimates biased. Nevertheless, like most research that rely on observational data, we are unable to fully rule out the existence of hidden confounders or other factors that can cause endogeneity in our setting. Thus, to ensure that our main findings are not biased due to these concerns, we take several steps. First, we rerun our analysis using an IV approach (see Section 5.1 for more details). Second, we repeat our analyses after making use of propensity score matching and creating balanced covariates (see Section 5.2 for more details). Both our IV and matching analyses yield results that are fairly similar to our main findings, indicating that our results are less likely to be biased due to potential issues discussed above. Third, we also perform formal tests and investigate a potential channel that explains the underlying causal mechanism behind our findings (see Section 4.2 for more details), which gives us further confidence about the validity of our main findings. However, we still emphasize the need to conduct a carefully designed Randomized Controlled Trial (RCT) in future research to further investigate the validity of our findings. Before future research does this, we acknowledge that some of our findings might be affected by factors that are unobservable to us, and hence, highlight that hospital administrators should primarily interpret our results at an association level (see Section 6 for further discussions). Finally, as part of our limitations section (see Section 6), we discuss that our results are based on findings in a single institution, and hence, generalization to other contexts (e.g., other hospitals) might require further consideration.

4. Results

We provide descriptive statistics and correlations in Table 1. The table displays weakly positive correlations between each of the task categories, except for the high complexity and resource-intensive tasks, which demonstrate a strongly positive correlation ($r = 0.749$). The stronger relationship between high complexity and resource-intensive tasks is expected, since highly complex tasks potentially involve a greater breadth of issues, the integration of a diverse set of activities, and more uncertainty. Thus, handling highly complex tasks typically require making use of many resources. To account for potential collinearity in our models, we residualized the resource-intensive task variable from high complexity tasks.

We started our analyses by characterizing when deviations occur in our context, answering the following question: do deviations from preferred assignments occur more during (1) high volume periods (versus low volume periods), (2) weekdays (versus weekends), and (3) earlier shifts (versus

Table 1 Descriptive Statistics and Correlations

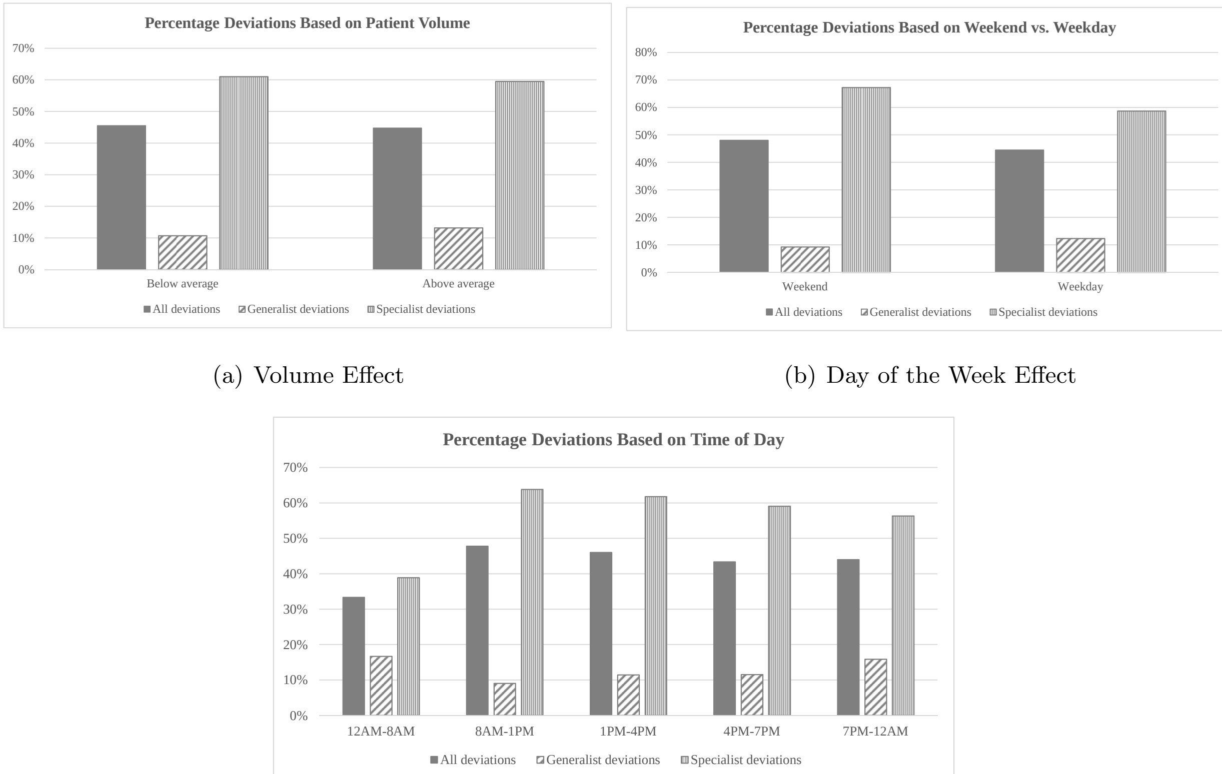
	Variable	Mean	s.d.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Preferred assignment	0.548	0.498	1															
2	Type of physician assigned (specialty)	3.078	2.022	-0.4444*	1														
3	Length of stay (days)	5.900	14.774	0.0273	0.0794*	1													
4	Total charges ($)	54,800	200,798	0.0313*	0.0943*	0.9317*	1												
5	Readmission, 30-day	0.065	0.246	-0.0148	-0.0144	0.0334*	0.0310*	1											
6	Adverse event	0.070	0.256	-0.0754*	0.1016*	0.0491*	0.0538*	-0.0287*	1										
7	Well-defined task	0.681	0.466	-0.0143	0.0790*	-0.0451*	-0.0308*	-0.1060*	0.0410*	1									
8	High complexity task	0.484	0.500	0.1788*	-0.024	0.0346*	0.0466*	-0.0879*	-0.1343*	0.2510*	1								
9	High contact task	0.357	0.479	0.1300*	0.1753*	0.0064	0.0390*	-0.0791*	-0.0375*	0.1168*	0.3286*	1							
10	Resource-intensive task	0.382	0.486	0.1792*	-0.0236	0.0262	0.0449*	-0.0900*	-0.1177*	0.1630*	0.7488*	0.5144*	1						
11	Patient volume	2.123	1.570	0.0157	-0.0450*	0.0105	0.0215	0.0111	0.0089	-0.0162	0.0157	-0.0335*	0.0122	1					
12	Patient age (years)	7.786	6.483	-0.1154*	0.1174*	-0.0314*	-0.0138	0.0966*	0.0415*	0.1117*	0.0051	0.0227	0.0207	0.015	1				
13	Male patient	0.474	0.499	-0.0242	-0.0175	-0.0032	-0.0028	-0.1052*	-0.0016	0.0116	0.0181	0.0215	0.0082	-0.0351*	-0.0615*	1			
14	Private insurance	0.303	0.460	-0.0382*	0.0591*	-0.0336*	-0.0267	0.0528*	0.0162	0.001	-0.0360*	0.0036	-0.0366*	0.0135	0.0061	0.0035	1		
15	CCI	1.537	1.030	0.0208	-0.0415*	0.0696*	0.0434*	0.0331*	-0.0199	-0.1277*	0.1142*	-0.0031	0.1370*	0.0267	0.0154	0.0028	-0.0263	1	
16	Flu season	0.590	0.492	0.0018	0.0269	0.014	0.0089	-0.0025	0.0095	0.005	-0.003	-0.0072	-0.0166	0.0369*	-0.0136	-0.021	-0.0107	-0.0108	1

Notes: CCI = Chronic Condition Indicator. $n = 4,729$. $^*p < 0.05$.

later shifts)? The descriptive findings are presented in Figure 1. Part (a) of this figure indicates that percentage deviations do not differ notably between high volume and low volume periods. Specifically, percentage of deviations both when a generalist is assigned (labeled as "generalist deviations") and when a specialist is assigned (labeled as "specialist deviations") are similar between periods with below average and above average patient volume. However, part (b) of Figure 1 reveals that deviations occur more during weekends than during weekdays, and that this is primarily associated with the fact that deviations when a specialist is assigned (while a generalist should have been assigned) is much higher during weekends than during weekdays. Finally, part (c) of Figure 1 shows that the highest and lowest percentage of deviations occur during morning shifts (8am-1pm) and after midnight shifts (12am-8am), respectively.

Next, we examined the effect of task type on preferred assignment, as shown in Table 2. The analysis also included professional type (generalist or specialist, as categorized across each of the eight specialties represented by the physicians in our sample) to further examine patterns in preferred assignment that relate specifically to professional role that may be due to knowledge differences as other research suggests (see, e.g., Atkinson et al. 2018, Shafritz et al. 2015, Crowston 1997, Galbraith and Galbraith 1977). Since the dependent variable is binary, we used a GLM model with binomial family and logit link, with standard errors clustered at the physician level. Average marginal effects (AME) are also shown in Table 2. The model includes controls for patient characteristics (i.e., various demographic and diagnostic variables discussed earlier) as well as year fixed effects (the full versions of our tables can be found in the appendix).

From Table 2, we observe the following. First, the type of physician assigned had a negative and statistically significant estimated coefficient. In addition, we observe that well-defined and resource-intensive tasks had a statistically insignificant coefficient. However, tasks that were high in complexity and contact had statistically significant results. High complexity tasks had a positive coefficient ($p < 0.001$) and AME results indicating a 14.8% increase in the likelihood of preferred assignment, or 27.0% when compared to the sample average of 54.8%. For high contact tasks, the results indicate a negative coefficient ($p < 0.001$), with an AME showing that high contact tasks



(a) Volume Effect

(b) Day of the Week Effect

(c) Time of the Day Effect

Figure 1 Percentage Deviations Based on Patient Volume, Day of the Week, and Time of the Day

Table 2 Task-Related Factors Associated with Preferred Assignment

Dependent Variable: Preferred Assignment	Coefficients	AME
Type of physician assigned (specialty)	-0.201*** (0.0496)	–
Well-defined task	-0.126 (0.100)	-0.0173
High complexity task	1.071*** (0.116)	0.1483
High contact task	-0.523*** (0.140)	-0.0718
Resource-intensive task	0.209 (0.154)	0.0494

Notes: Generalized linear model results reported (binomial family, logit link). Standard errors are in parentheses. Model is adjusted by patient characteristics and patient volume, and clustered by physician assigned. Includes year fixed effects. *** $p < 0.001$. AME = Average marginal effect.

decrease the probability of preferred assignment by 7.1%, or 12.9% when compared to the sample average of 54.8%.

In Table 3, we report the performance implications of preferred assignment and task characteristics as a precursor to testing our hypotheses. Results for operational efficiency (length of stay) and cost (total charges) are reported in M1 and M2 using GLM models with a gamma family and logistic link, while quality outcomes (readmission and adverse event rates) are reported in M3 and M4 using GLM models with binomial family and logit link. All models include adjustments for patient characteristics and year fixed effects, with standard errors clustered by physician. As M1 and M2 show, preferred assignment has a statistically significant effect on costs ($p < 0.001$), specifically in increasing the length of stay and total charges. M3 and M4 indicate no statistically significant relationship between preferred assignment and quality outcomes. With regard to task categories, M1 and M2 show that well-defined tasks do not have a statistically significant effect on length of stay and cost outcomes, though have a statistically significant negative effect on readmission rates ($p < 0.001$) and statistically significant positive effect on adverse event rates ($p < 0.001$), as displayed in M3 and M4. Thus, well-defined tasks have a mixed effect on quality outcomes, demonstrating lower readmission occurrences yet higher instances of adverse events.

For high complexity tasks, M1 and M2 in Table 3 indicate no statistically significant effects on length of stay and cost outcomes. However, M3 and M4 show that high complexity tasks, compared to those with low complexity, have readmission occurrences that were significantly lower ($p < 0.05$), and adverse event instances that were also significantly lower ($p < 0.001$). Considering M1 and M2, we also see that tasks requiring high (versus low) contact and those that were resource-intensive (versus those with low resource requirements) had significantly longer length of stay and higher total charges. M3 and M4 show no statistically significant effects on either of the quality outcomes for high contact tasks, yet statistically significant declines in 30-day readmission for resource intensive tasks ($p < 0.05$). Table 4 provides a summary of the results in Table 3 and highlights our main findings.

To further examine the results from Tables 3 and 4, specifically the combined effect of task characteristics and preferred assignment on performance, we tested interaction effects. The results are shown in Table EC.1 in the Appendix. Our first hypothesis (Hypothesis 1) is that following the preferred assignment will yield higher performance for tasks that are well-defined. The results in Table EC.1 (see M5 and M10 there) show that the *well-defined* $\times$ *preferred assignment* interaction has a negative coefficient for both operational efficiency and cost outcomes ($p < 0.001$ for length of stay and total charges). However, the interaction is statistically insignificant for both quality outcomes, as shown in M15 and M20. Figures 2(a) and 2(b) graphically capture the marginal effects of preferred assignment and well-defined tasks, indicating that Hypothesis 1 is not supported by the significant interactions pertaining to length of stay and total charges outcomes. Specifically, comparing preferred and deviating assignments, we observe that preferred assignments are associated with an increase of 0.304 days and \$6,160 in length of stay and total charges, respectively, for well-defined tasks. However, the graphs also demonstrate that (1) well-defined tasks have lower length of stay and total charges compared to poorly-defined tasks when assignment is preferred; and (2) although preferred assignment, compared to deviating assignment, tends to increase both length of stay and total charges, the effect is greater for poorly-defined tasks. These findings indi-

Table 3 Performance Implications of Preferred Assignment and Task Categories

Dependent Variable:	Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse event
Variables	M1	M2	M3	M4
Preferred assignment	0.144***	0.181***	0.155	-0.0559
	(0.0344)	(0.0355)	(0.166)	(0.180)
Type of physician assigned (specialty)	0.0210	0.0293	-0.01105	0.156**
	(0.0140)	(0.0155)	(0.0451)	(0.0496)
Well-defined	0.00253	-0.00254	-0.749***	0.618***
	(0.0312)	(0.0321)	(0.136)	(0.157)
High complexity	0.00254	0.0314	-0.411*	-1.195***
	(0.0345)	(0.0356)	(0.164)	(0.175)
High contact	0.118**	0.233***	-0.330	-0.146
	(0.0431)	(0.0444)	(0.197)	(0.208)
Resource-intensive	0.102*	0.139**	-0.525*	-0.439
	(0.0480)	(0.0496)	(0.236)	(0.270)
Constant	1.906***	10.58***	-3.096***	-2.648***
	(0.213)	(0.225)	(1.119)	(0.930)
N	4,729	4,729	4,729	4,729

Notes: Generalized linear model results reported (M1-M2: gamma family, logistic link; M3-M4: binomial family, logit link). Standard errors are in parentheses. Model is adjusted by patient characteristics and patient volume, and clustered by physician assigned. Includes year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4 Summary of Performance Implications Across Task Categories

Dependent Variable:	Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse event
Well-defined task	0	0	-	+
High complexity task	0	0	-	-
High contact task	+	+	0	0
Resource-intensive task	+	+	-	0

Notes: + = positive effect; - = negative effect; 0 = no effect.

cate that following the preferred assignment is more detrimental for poorly defined tasks (increase in length of stay by 2.00 days and in total charges by $19,983), supporting the notion that deviations in assignment benefits operational efficiency and cost outcomes more for poorly-defined tasks than for well-defined tasks.

Hypothesis 2 postulates that following the preferred assignment will result in worse performance for tasks with high complexity. In M6 and M11 (Table EC.1 in the Appendix), we see no statistically significant effect related to the interaction *high complexity* × *preferred assignment* on either operational efficiency or cost. Similarly, M16 and M21 indicate no statistically significant effect on either of the two dimensions of quality as it relates to high complexity and preferred assignment.

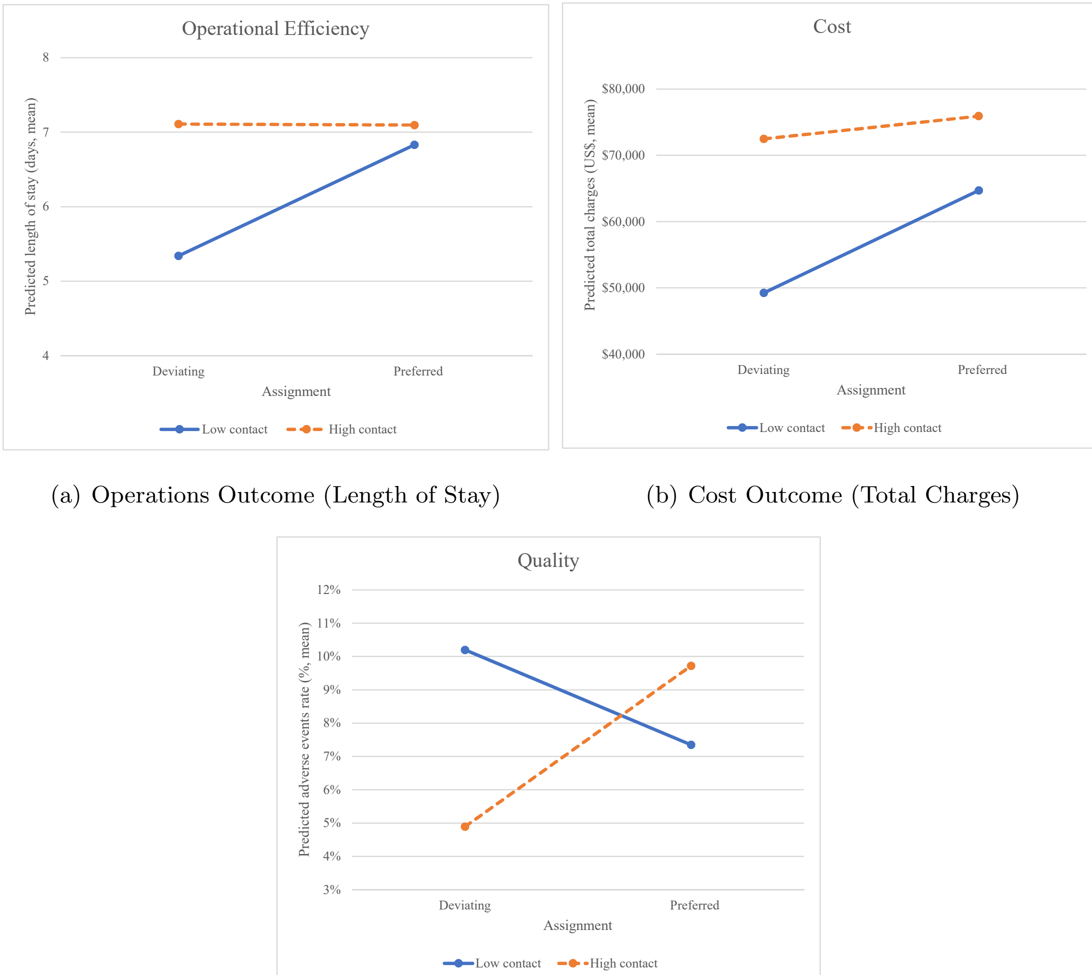
Hypothesis 3 states that preferred assignment results in worse performance when tasks involve high contact. M7 and M12 in Table EC.1 (see the Appendix) show that both length of stay and

(a) Operations Outcome (Length of Stay) (b) Cost Outcome (Total Charges)

Figure 2 Predicted Performance Outcomes Resulting from Interaction of Preferred Assignment and Well-Defined Tasks

total charges have a statistically significant negative coefficient ($p < 0.001$ and $p < 0.01$) for the *high contact* × *preferred assignment* interaction. In terms of quality outcomes, M22 shows a statistically significant positive effect ($p < 0.001$) on adverse events, indicating that for high contact tasks preferred assignments have an increased incidence of adverse events. However, our results show statistically insignificant effects in the other quality outcome, 30-day readmission, as shown in M17. To further evaluate these effects, Figures 3(a) and 3(b) demonstrate the predicted length of stay and cost outcomes associated with *high contact* × *preferred assignment*. These graphs show partial support for our hypothesis, in that compared to deviations in assignment, preferred assignment reduces length of stay by 0.013 days but increases total charges by $3,432. However, the figures also demonstrate a steeper slope for low contact tasks, indicating that deviation is likely more beneficial for low compared to high contact tasks. In addition, from Figure 3(c), we observe that preferred assignment for high contact tasks increases the likelihood of adverse events by 4.83%, compared to situations when assignment deviates. Thus, our findings are somewhat mixed for outcomes related to high contact tasks with preferred assignment: while operational efficiency improves, cost and quality (in terms of adverse events) degrade.

Finally, Hypothesis 4 suggests that following the preferred assignment improves performance for tasks that are resource-intensive. In Table EC.1 (see the Appendix), M8 and M13 show statistically significant negative effects in support of this hypothesis. We see no statistically significant results, however, for the quality measures in M18 and M23. In examining this interaction (*resource-intensive* × *preferred assignment*), Figures 4(a) and 4(b) provide graphs illustrating the moderating effect, which show that in comparison to deviations in assignment, preferred assignment involves lower length of stay by approximately 4.4 days and lower total charges by $24,030 when tasks have high resource intensity.



(a) Operations Outcome (Length of Stay)

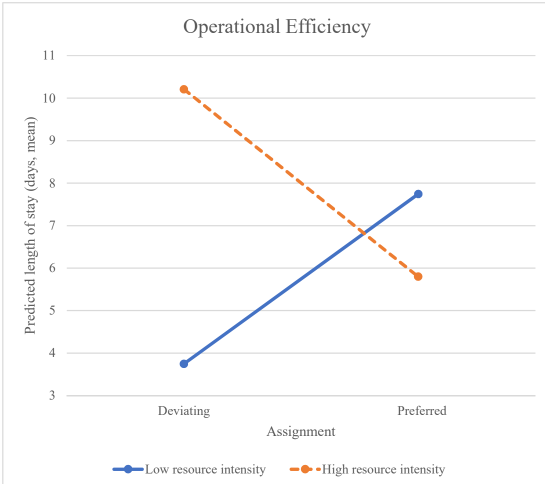
(b) Cost Outcome (Total Charges)

(c) Quality Outcome (Adverse Events)

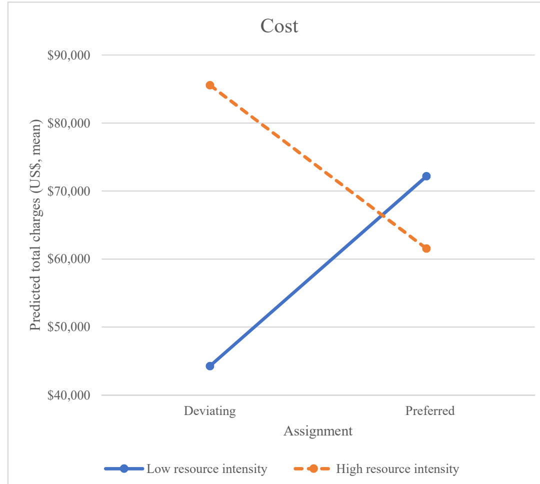
Figure 3 Predicted Performance Outcomes Resulting from Interaction of Preferred Assignment and High Contact Tasks

4.1. Summary of Hypothesis Tests

The summary of our results regarding our hypothesis tests is provided in Table 5. The table shows that the majority of our hypotheses cannot be rejected, especially when we consider all three sets of performance outcomes (operational efficiency, cost, and quality). However, the table indicates that we can reject our first hypothesis (Hypothesis 1) with respect to operational efficiency and total cost outcomes, implying that following preferred assignment for well-defined tasks on average worsens these performance outcomes. This runs counter to our proposed hypothesis; one reason may be that, since these tasks are more straight-forward, any professional (generalist or specialist) has the necessary information and skills to be assigned. Thus, it may be better to assign either role in light of other organizational conditions (e.g., staff availability, personal experience dealing with these tasks, etc.).



(a) Operations Outcome (Length of Stay)



(b) Cost Outcome (Total Charges)

Figure 4 Predicted Performance Outcomes Resulting from Interaction of Preferred Assignment and Resource-Intensive Tasks

Table 5 Summary of Hypothesis Tests

Hypothesis #	Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission rate, 30-day	Adverse event
1	F	F	–	–
2	–	–	–	–
3	F	T	–	T
4	T	T	–	–

Notes: T = true; F = false; – = insignificant results.

We can also reject Hypothesis 3 with respect to operational efficiency. Specifically, deviations in assignment for high contact tasks produce worse outcomes on average in terms of the patients' length of stay. A possible reason for this is that high contact tasks involve patients who more frequently interact with their assigned physician, and could therefore extend their hospitalization time, since they are more likely to communicate issues; a deviation in assignment from the preferred professional could prolong the clinicians response time and the ability to manage these issues, if the professional is not as experienced or versed in dealing with the patient's concerns. On the other hand, our results indicate that Hypothesis 3 cannot be rejected with respect to total cost and adverse event outcomes. The mixed results regarding Hypothesis 3 may reflect the nature of high contact tasks, in which patients may have more interactions with their assigned physician.

Finally, we observe that Hypothesis 4 cannot be rejected with respect to operational efficiency and total cost outcomes. Taken together, these results suggest that providing physicians with the flexibility to deviate from preferred assignments is beneficial in (a) well-defined tasks (improving operational efficiency and costs), and (b) high contact tasks (improving costs and adverse events, though at the expense of lower operational efficiency). For other tasks, preferred assignments should be enforced, as deviations either negatively impact performance or yield no tangible benefits.

4.2. Mechanism: Mediation Analysis

To shed light on a potential mechanism that can be behind our main findings, we now make use of mediation analysis. Specifically, we examine whether and how the use of advanced imaging during hospitalization can mediate the relationship between assignment matching the preferred and the four outcomes of interest (length of stay, total charges, 30-day readmission, and adverse events). To this end, we first identify hospitalizations in which resources such as MRIs, CT scans, or nuclear radiology are used at least once (approximately 25% of the hospitalizations in our sample). We then test the mediating impact of utilizing such resources and conduct separate tests for each task type. To do these, we employ a bootstrapping approach to run simple mediation models (see, e.g., Valeri and VanderWeele 2013), and estimate the direct and indirect effects of preferred assignment (by task type via the advanced imaging mediator) on each of the four outcome variables. Our model specifications include the following: since length of stay is a count variable, we use Poisson regression; since total charges represents a continuous variable, we use linear regression, and since 30-day readmission and adverse events are both binary outcomes, we use logistic regression.

The main results of our mediation analysis are shown in Table EC.2 (see the Appendix) for each task type: well-defined tasks ($n = 3,222$), high complexity tasks ($n = 2,289$), high contact tasks ($n = 1,686$), and resource-intensive tasks ($n = 1,806$). With regard to well-known tasks, we find that advanced imaging has a mediating effect for length of stay and total charges. Employing bootstrapping analysis with 1,000 iterations, preferred assignment for high contact tasks shows a significant effect on advanced imaging ($B = -0.291$, $se = 0.079$, $p < 0.01$), which in turn has a significant effect on length of stay ($B = 0.996$, $se = 0.015$, $p < 0.001$) and total charges ($B = 68,046$, $se = 9263$, $p < 0.001$). With regard to high contact tasks, we find mediating effects for length of stay and total charges as well. Specifically, we find a significant effect of preferred assignment on advanced imaging ($B = -0.503$, $se = 0.129$, $p < 0.001$), which in turn has a significant effect on length of stay ($B = 1.138$, $se = 0.021$, $p < 0.001$) and total charges ($B = 106,792$, $se = 16,719$, $p < 0.001$). In terms of resource-intensive tasks, preferred assignment demonstrates a significant effect on advanced imaging ($B = -0.294$, $se = 0.120$, $p < 0.05$) which subsequently has a significant effect on total charges ($B = 87,622$, $se = 15,401$, $p < 0.001$). We do not observe mediation effects for highly complex tasks, as these models have insignificant NIE values.

To gain a deeper understanding, we also examine whether the mediation effect of utilizing advanced imaging resources depends on the type of the provider that is assigned. This allows us to generate more insights into the mediation effect of use of advanced imaging, especially when we note that (a) specialist and generalists tend to use advanced imaging at different rates (even after adjusting for other variables such as patient conditions, comorbidities, etc.), and (b) whether or not there is a deviation in assignment can impact use of advanced imaging, since it affects the type of physician that is assigned. Table EC.3 (see the Appendix) shows the result of running separate models for hospitalizations in which a generalist is assigned ($n = 1,477$) and those in which a specialist is assigned ($n = 3,252$), both according to the guideline (i.e., what the preferred assignment indicates). Generalist assignment according to guideline shows a significant effect on advanced imaging ($B = -1.502$, $se = 0.232$, $p < 0.001$), which in turn has a significant effect on length of stay ($B = 1.007$, $se = 0.025$, $p < 0.001$) and total charges ($B = 53,474$, $se = 4,663$, $p < 0.001$), indicating meditating effects for these outcomes. However, the NIE is not significant for 30-day readmission and the occurrence of adverse events, and hence, we do not observe mediating effects on these outcomes.

We follow similar procedures for analyzing the hospitalizations in which a specialist is assigned. Specialists assignment according to guidelines has a significant effect on advanced imaging ($B = 0.544$, $se = 0.103$, $p < 0.001$), which subsequently has a significant effect on length of stay ($B = 0.992$, $se = 0.014$, $p < 0.001$) and total charges ($B = 81,019$, $se = 9,453$, $p < 0.001$), indicating meditating effects for these outcomes. However, we again do not observe significant mediating effects for the 30-day readmission and adverse event outcomes (NIE is not significant in these cases).

In summary, our mediation analysis shows that advanced imaging mediates the relationship between preferred assignment and total charges for resource-intensive tasks. In addition, we find that advanced imaging partially mediates the relationship between following the preferred assignment and outcome variables such as length of stay and total charges, especially for high contact tasks. The mediating effect of use of advanced imaging also depends on whether the preferred assignment suggests putting a specialist or a generalist in charge. This heterogeneity is, however, to some extent expected, given that the use of advanced imaging is not similar across these two types of providers.

In interpreting our results, we caution that our conclusions might be biased due to potential mediator-outcome confounding (see, e.g., VanderWeele 2015) or other concerns related to the use of “bad controls” (see, e.g., Angrist and Pischke 2009). Our various investigations, however, suggest that these concerns in our setting are likely mitigated. This is partially because we are using a exposure-mediator interaction method (Valeri and VanderWeele 2013, VanderWeele 2015), and concerns about the mediator being endogenous is a bit less prominent in this method than traditional methods. Furthermore, as Valeri and VanderWeele (2013) state, to control for mediator-outcome confounding “the investigator must adjust for common causes of the mediator and the outcome.” In our setting, we have adjusted for many variables that could be common causes of the mediator and outcomes (see, e.g., the list of variables in Section 3.3). For these and other reasons, in our equations the correlation between the error terms and the mediator are fairly weak, suggesting that the mediator is likely exogenous. More importantly, use of sensitivity analysis (see, e.g., Section 5.6 in VanderWeele 2015) suggest that even if our estimates are biased due to these concerns,

Table 6 Predicted Impact of Different Counterfactual Scenarios (Compared to Current Practice) Due to Adhering to Preferred Assignment

Scenario	LOS (days)	Charges ($)	Readmission (%)	Adverse Events (%)
Scenario 1 (all tasks)	-0.046	-$538	–	0.01%
Scenario 2 (well-defined tasks only)	0.466	$8,474	–	–
Scenario 3 (high complexity tasks only)	–	–	–	–
Scenario 4 (high contact tasks only)	-0.072	$2,178	–	4.14%
Scenario 5 (resource-intensive tasks only)	-2.252	-$23,566	–	–

Notes: – = insignificant results.

our conclusions (e.g., directional effects) likely remain fairly robust when relevant adjustments are made. Nevertheless, we cannot fully rule out the potential impact of these important concerns. Thus, we believe our findings should mainly be interpreted in terms of correlations (or at best "suggestive" causations) that need to be further investigated in future research (e.g., through rigorous experiments).

4.3. Providing Recommendations Using Cost-Effectiveness Analyses

To provide clear recommendations for hospital administrators, if they choose to develop more formalized guidelines around preferred assignments, we now consider five counterfactual policies corresponding to imposing the guideline fully or partially. Specifically, in Table 6 we summarize the predicted impact of adhering to preferred assignment for all task types (Scenario 1), well-defined tasks only (Scenario 2), high complexity tasks only (Scenario 3), high contact tasks only (Scenario 4), and resource-intensive tasks only (Scenario 5).

Our results show that among these scenarios, Scenario 5 (imposing the guideline only for resource-intensive tasks) should be viewed as best followed by Scenario 1 (imposing the guideline for all task types). However, we find that other scenarios should be avoided as they can degrade performance compared to the current practice. In addition, we find that there is a tradeoff in Scenario 1: if it is pursued by hospital administrators, operational efficiency (LOS) and costs (total charges) would improve but quality (adverse events) would worsen.

To assist hospital administrators in understanding the underlying tradeoff in cost versus quality inherent in implementing Scenario 1, we make use of cost-effectiveness analysis. In particular, in Figure 5, we depict the region in which pursuing Scenario 1 is cost-effective. In this figure, we make use of the widely-accepted Willingness-To-Pay (WTP) rate of $100,000 per Quality-Adjusted Life Years (QALY). The figure indicates that at this WTP rate, imposing Scenario 1 is cost-effective unless the impact of the increase in adverse events on QALY is high (specifically, higher than about 0.53 QALYs per 1% adverse events).[7] Thus, Scenario 1 should be viewed as a cost-effective alternative to Scenario 5. However, as noted earlier, our results indicate that Scenario 5 is the best

[7] Since our partner hospital is a children's hospital, we note that adverse events can be more consequential (i.e., have long-term and lasting effects on patients) than in non-children hospitals. Yet, it is unlikely that the average impact of adverse events on QALY can reach this high level.

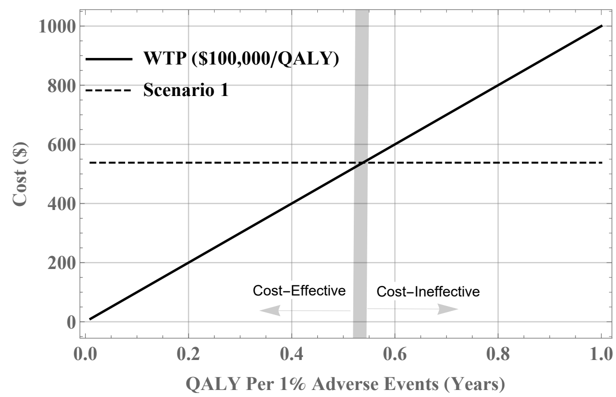


Figure 5 Cost-Effectiveness of Scenario 1 (Imposing the Guideline for All Task Types).

scenario among the five counterfactual ones considered, as it yields tangible benefits on operational efficiency and costs measures without any negative impact on metrics related to quality.

5. Robustness Checks

We now perform various robustness checks to ensure that our main findings are not affected by endogeneity issues, the measurement used to define consensus opinion, the set of controls involved, or other model specifications. To this end, we re-run our analyses by making use of (a) instrumental variable (IV) analysis, (b) matching, (c) alternative approaches for measuring consensus opinion, (d) additional control variables, and (e) adjustments that allow us test the potential impact of inherent correlations between our outcome variables.

5.1. Instrumental Variable (IV) Analysis

To address potential concerns with endogeneity that may be associated with unobservable variables in our analysis, we used an instrumental variable (IV) approach with specifications comparable to our original GLM models. Specifically, we fit GLM models using 2-stage nonlinear least squares. The details of our IV analysis, which we used as one way of checking the robustness of our main findings presented in the previous sections, are provided in the Appendix. Overall, our results gives us confidence that our main findings are relatively robust, and not biased due to potential endogeneity issues.

5.2. Matching Analysis

To further corroborate our findings, we re-ran our analyses using a matched sample. Specifically, we made use of the 1-nearest neighbor propensity score matching approach to first balance the covariates in our treatment (i.e., assignment matching preferred/guideline) and control (i.e., assignment deviating from preferred/guideline) groups. We carried out models for each outcome measure, specifying a caliper of 3 for our baseline analysis (we also varied this caliber to further ensure robustness). The main results are provided in Tables 7.[8] As can be seen from Table 7, the results

[8] Further details, including the main matching variables and balance of covariates, are presented in Table EC.4.

Table 7 Summary of Performance Implications Across Task Categories, Using the 1-Nearest Neighbor Propensity Score Matching Approach

Dependent Variable:	**Operational Efficiency**	**Cost**	**Quality**	
	Length of stay (days)	**Total charges ($)**	**Readmission, 30-day**	**Adverse event**
Well-defined task	-	-	-	0
High complexity task	+	0	-	-
High contact task	-	0	-	0
Resource-intensive task	0	0	+	-

Notes: + = positive effect; - = negative effect; 0 = no effect.

support our prior findings, with the exception of high contact tasks with respect to the length of stay outcome, and resource-intensive tasks with respect to the 30-day readmission outcome. The fact that our findings did not change notably from our prior analysis gives us further confidence about the robustness our results. Finally, Table EC.4 presents the standard differences and variance ratios for the main patient characteristics used for matching. For each of the covariates, comparing across the treatment and control groups, the standard difference is very close to zero and the variance ratios approximated one. This suggests that our approach has been effective in creating nearly balanced covariates, and hence, the fact that we observe similar results with and without matching should not be attributed to a potential inefficiency in our matching approach in creating matched samples.

5.3. Measuring Consensus Opinion

We used physicians' consensus opinion from the survey we administered to determine the preferred assignment. Consensus occurred when *at least* 50% of the physicians surveyed agreed that a generalist or a specialist should be assigned to patients with a particular diagnosis. Using a threshold of 50% as the minimum level is consistent with the literature. For example, similar to our approach, in de Rooij et al. (2020), a consensus is considered when at least 50% of the votes indicate a specific outcome. Likewise, in studies such as Lynch et al. (2008), Rahbari et al. (2009), Waldman et al. (2011), van de Velde et al. (2014), Miskovic et al. (2015), and Nieuwenhuys et al. (2016), lack of consensus is defined as when the votes from respondents fall below a 50% agreement. Nonetheless, to test the robustness of our results to this definition of consensus opinion, we used two different approaches.

First, we measured percentage agreement for each diagnosis in our survey and ran each model separately by altering the threshold on the minimum percent agreement to 45% and 55% instead of 50%. Tables EC.5 and EC.6 (see the Appendix) show that the results are consistent with our original findings.

In our second approach, we calculated a reliability coefficient which is part of the derivation of *Fleiss' inter-rater reliability measure*, kappa (Fleiss 1971):

$$P_i = \frac{1}{n(n-1)} \sum_{j=1}^{k} n_{ij}(n_{ij}-1),$$

where n is the total number of raters (medical expert respondents), i is the item being rated (diagnoses), j represents the category selected by the raters (generalist or specialist assigned), k is the total number of categories ($k=2$ in our setting), and n_{ij} is the number of raters who chose category j for item i. We used this approach as an alternative, because P_i can be used with small or variable sample sizes, and in calculating it agreement is weighted by the number of expert raters. Furthermore, this approach considers the fact that agreements by chance are unlikely, since the expert raters likely have previous knowledge about the items in the survey. P_i can range from 0 (perfect disagreement) to 1 (perfect agreement). We chose a threshold of 0.5 on P_i (as is common in the literature) to define consensus opinion for each medical diagnosis in the survey. Table EC.7 (see the Appendix) displays similar results to our earlier reported findings, supporting the robustness of our results to our original measure for consensus opinion. Finally, we observed that changing the threshold of 0.5 on P_i does not significantly change our findings.

5.4. Additional Control Variables

To further test the robustness of our results, and check for the possible impact of other variables related to physician workflow, we incorporated two additional control variables. First, we created a variable for *weekend service*, which represents when a patient was served on a Saturday or Sunday when staffing is more limited versus another time during the week. Second, we included a variable for *shift change*, to represent the times in which providers would either just arrive to their shift or just leave their shift. At our partner hospital, shift changes are at 8am and 5pm so we included the 10-minute window around those times and our variable captures whether patients were served during those periods. In both instances, our results show robustness, matching our prior findings. Specifically, we observe that adding these control variables does not significantly impact our findings, indicating that our set of original control variables capture the important variations among different hospitalizations in our data set.

5.5. Bonferroni Correction

We also assessed the robustness of our results to the assumption that our outcome variables are independent. If this assumption is violated, the chance of incorrectly rejecting a null hypothesis (i.e., making a Type I error) in our setting increases. In particular, our hypothesis tests might have yielded incorrect results purely by chance, since we conduct multiple comparisons on different outcomes. To address this concern, we first made use of correlation analysis to directly test the level of correlation between our outcome variables. We observed that the only considerable level of correlation is between total charges and length of stay. This is expected, given that a longer length of stay almost always involves additional expenditures due to use of additional resources such as extra tests, staff time, and higher bed usage, among others. Regardless, to fully examine to robustness of our results to the potential correlation among all the outcome variables, we then applied a very conservative approach. Namely, we used Bonferroni correction, by simply adjusting the significant p-values, such that a minimum threshold of 0.0125 (or 0.05 divided by 4 outcome variables) would indicate significance. In applying this conservative approach, we still attained significant or marginally significant results, and observed that our original results hold. This gives us further confidence about the robustness of our findings to the assumption of outcome independence. In addition to this conservative Bonferroni adjustment, we also considered using other, less conservative approaches, including Hochberg, Hommel, Holm, and Benjamini Hochberg and Yekutieli procedures. These approaches vary in their level of conservativeness. The Bonferroni test is, however, the most conservative (Benjamini and Hochberg 1995) one and gives us the maximum confidence about the robustness of our results.

6. Limitations

Our study has some limitations. First, we performed this analysis at a single institution in the healthcare sector. Future work can conduct a similar analysis in other hospitals, as well as across other sectors that rely on professional expertise that may overlap. Another limitation of this study is that the implementation of preferred assignment guidelines did not formally occur in our analysis. Specifically, we built our understanding of preferred assignment on prior work explaining how they are typically a product of consensus opinion (Brivot 2011). Thus, the way the majority of physicians in our sample viewed assignments was retrospectively chosen as the main factor determining the preferred practice. In the ideal study, which future research in this area could attempt, preferred practices would be formally made known as guidelines to professionals, and analysis of performance could occur using prospective data. In addition, we controlled for the nature of the work being performed by generalist and specialist physicians, namely the nature of the patients' diagnoses in our data set. However, we may be missing aspects of the task that could impact performance outcomes. For example, 30-day readmissions were not adjusted to account for whether a diagnosis involved a scheduled future hospitalization due to the nature of the condition and follow-up treatment required. Finally, while our various robustness checks indicate that our main findings are fairly robust and not affected by potential endogeneity issues (that can make our result biased), how consensus opinion is measured, the set of controls included, or other model specifications, deriving causal conclusions from an observational data set like ours can still be subject to various errors. Thus, one needs to be cautious in making such conclusions. More importantly, given the significant implications that our findings can have for hospitals, we hope future research can conduct an appropriate RCT, which can further test the validity of the evidence we establish.

7. Concluding Remarks

Our work is motivated by the contrast between (a) operations management and organizational theories that promote standardization and reduction of deviations from preferred assignments, and

(b) decision theories and the conventional wisdom that suggest professionals should be given the opportunity to deviate as needed. Our results take a *task-type* view of this contrast by generating insights into specific task types for which deviations can improve performance.

Our study shows that providing physicians with the flexibility to deviate from preferred assignments can improve performance when (a) patient needs are well-defined (improving operational efficiency and cost), or (b) serving the patient requires high contact (improving costs and adverse events, though at the expense of lower operational efficiency). For other task types, we find that hospital administrators should enforce preferred assignments. In addition, our mediation analysis aimed at understanding the mechanism behind our findings indicate that use of advanced imaging (e.g., MRIs, CT scans, or nuclear radiology) plays a significant role in how deviations impact performance outcomes. This indicates that, at least in our partner hospital, hospital administrators should better regulate use of advanced imaging. We expect that better understanding the underlying differences between specialists and generalists in using resources such as MRIs, CT scans, and nuclear radiology, and consequently providing appropriate training programs can go a long way.

Our results also provide evidence for a *no free lunch theorem*: while for some task types deviations from preferred assignments are beneficial from some aspects, they are simultaneously detrimental from other aspects. Hence, in contrast with the finding that enforcing preferred assignments is the dominant strategy for some tasks types (e.g., resource-intensive tasks), permitting deviations is typically not dominantly the better option (regardless of the task type). Furthermore, our results indicate that there may be an intriguing interplay between professional status and deviations from preferred assignments in practice, which is worthy of future research as this potential assertion of power in deviating from standard practice changes division of labor and can have negative consequences on some outcomes.

Finally, our results show that there might be environmental conditions under which deviations occur more frequently. In particular, we find that such deviations occur more during weekends than weekdays, and during morning shifts (8am-1pm) than other shifts. However, we observe that deviations occur similarly during high congestion (busy) and low congestion (less busy) periods, suggesting that congestion might not be an influential environmental factor. Future research can further investigate these issues and provide more insights into changes in environmental conditions that can cause an increase in deviations from preferred assignments.

Acknowledgments

We are grateful to all of those who provided invaluable feedback on this work, especially Vern Glaser, Trish Reay, Peer Fiss, John Romley, Paul Adler, Sara Singer, George Atkinson, Ankur Pandya, Nancy Kane, Jennifer Candipan, and Ben Sommers. Statistical support was provided by data science specialist, Simo Goshev, at the Institute for Quantitative Social Science, Harvard University. The second author is especially grateful for the assistance he received from his undergraduate RA, Sumhith Aradhyula. This study was partially supported by the first author's National Science Foundation grant (award #1459996) and the second author's National Science Foundation grant (award #1562645).

References

Aime, Federico, Scott Johnson, Jason W. Ridge, Aaron D. Hill. 2010. The routine may be stable but the advantage is not: Competitive implications of key employee mobility. *Strategic Management Journal* **31**(1) 75–87.

Angrist, Joshua D, Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Armony, Mor, Amy R. Ward. 2010. Fair dynamic routing in large-scale heterogeneous-server systems. *Operations Research* **58**(3) 624–637.

Atkinson, Mariam Krikorian, Mark A. Schuster, Jeremy Y. Feng, Temilola Akinola, Kathryn L. Clark, Benjamin D. Sommers. 2018. Adverse Events and Patient Outcomes Among Hospitalized Children Cared for by General Pediatricians vs Hospitalists. *JAMA Network Open* **1**(8) e185658–e185658.

Auerbach, A. D., M. B. Hamel, R. B. Davis, A. F. Connors Jr, C. Regueiro, N. Desbiens, L. Goldman, R. M. Califf, N. V. Dawson, N. Wenger. 2000. Resource use and survival of patients hospitalized with congestive heart failure: Differences in care by specialty of the attending physician. *Annals of Internal Medicine* **123**(3) 191–200.

Benjamini, Y., Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**(1) 289—-300.

Berry, Jay G., Sara L. Toomey, Alan M. Zaslavsky, Ashish K. Jha, Mari M. Nakamura, David J. Klein, Jeremy Y. Feng, Shanna Shulman, Vincent W. Chiang, William Kaplan, Matt Hall, Mark A. Schuster. 2013. Pediatric Readmission Prevalence and Variability Across Hospitals. *The Journal of the American Medical Association* **309**(4) 372–380.

Bodenreider, Olivier, Barry Smith, Anita Burgun. 2004. The ontology-epistemology divide: A case study in medical terminology. IOS Press, 185.

Boloori, Alireza, Soroush Saghafian, Harini A Chakkera, Curtiss B Cook. 2020. Data-driven management of post-transplant medications: An ambiguous partially observable Markov decision process approach. *Manufacturing & Service Operations Management* **22**(5) 1066–1087.

Boom, N. K., D. S. Lee, J. V. Tu. 2012. Comparison of processes of care and clinical outcomes for patients newly hospitalized for heart failure attended by different physician specialists. *American Heart Journal* **163**(2) 252–259.

Briscoe, Forrest, Michelle Rogan. 2015. Coordinating complex work: Knowledge networks, partner departures, and client relationship performance in a law firm. *Management Science* **62**(8) 2392–2411.

Brivot, Marion. 2011. Controls of knowledge production, sharing and use in bureaucratized professional service firms. *Organization Studies* **32**(4) 498–508.

Champy, Florent. 2009. *12. La Culture Professionnelle des Architectes*. La Découverte, 152—162.

Chase, Richard B., David A. Tansik. 1983. The customer contact model for organization design. *Management Science* **29**(9) 1037–1050.

Cohen, I. Glenn. 2013. Rationing legal services. *Journal of Legal Analysis* **5**(1) 221–307.

Crowston, Kevin. 1991. *Towards a Coordination Cookbook: Recipes for Multi-Agent Action*. MIT Sloan School of Management.

Crowston, Kevin. 1997. A coordination theory approach to organizational process design. *Organization Science* **8**(2) 157–175.

Currie, Graeme, Leroy White. 2012. Inter-professional barriers and knowledge brokering in an organizational context: The case of healthcare. *Organization Studies* **33**(10) 1333–1361.

Cyert, Richard M., James G. March. 1963. *A Behavioral Theory of the Firm*. Blackwell Publishers.

D'Adderio, Luciana. 2014. The replication dilemma unravelled: How organizations enact multiple goals in routine transfer. *Organization Science* **25**(5) 1325–1350.

Danet, Brenda. 1981. *Client-Organization Relationships*, vol. 2. Oxford University Press.

de Rooij, Maarten, Bas Israël, Marcia Tummers, Hashim U Ahmed, Tristan Barrett, Francesco Giganti, Bernd Hamm, Vibeke Løgager, Anwar Padhani, Valeria Panebianco, et al. 2020. ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training. *European Radiology* **30**(10) 5404–5416.

Dillman, Don A. 2011. *Mail and Internet surveys: The tailored design method–2007 Update with new Internet, visual, and mixed-mode guide*. John Wiley & Sons.

Engel, Gloria v. 1969. The effect of bureaucracy on the professional autonomy of the physician. *Journal of Health and Social Behavior* **10**(1) 30–41.

Feudtner, Chris, James Feinstein, Wenjun Zhong, Matt Hall, Dingwei Dai. 2014. Pediatric complex chronic conditions classification system version 2: Updated for icd-10 and complex medical technology dependence and transplantation. *BMC Pediatrics* **14**(1) 199.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* **76**(5) 378.

Galbraith, Jay, Jay R. Galbraith. 1977. *Organization Design*. Addison-Wesley.

Grant, Robert M. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal* **17**(S2) 109–122.

Greenfield, S., E. C. Nelson, M. Zubkoff, W. Manning, W. Rogers, R. L. Kravitz, A. Keller, A. R. Tarlov, J. E. Ware. 1992. Variations in resource utilization among medical specialties and systems of care: Results from the medical outcomes study. *JAMA* **276**(12) 1624–1630.

Harrold, T. S., L. R.and Field, J. H. Gurwitz. 1999. Knowledge, patterns of care, and outcomes of care for generalists and specialists. *Journal of General Internal Medicine* **144**(8) 499–511.

Hopp, W.J., S.M.R. Iravani, G.Y. Yuen. 2007. Operations systems with discretionary task completion. *Management Science* **53**(1) 61–77.

Ibanez, Maria R., Jonathan R. Clark, Robert S. Huckman, Bradley R. Staats. 2017. Discretionary task ordering: Queue management in radiological services. *Management Science* **64**(9) 4389–4407.

Jue, J. J., S. Cunningham, K. Lohr, P. Shekelle, R. Shiffman, C. Robbins, M. Nix, V. Coates, K. Schoelles. 2019. Developing and testing the agency for healthcare research and quality's national guideline clearinghouse extent of adherence to trustworthy standards (neats) instrument. *Annals of Internal Medicine* **170**(7) 480—-487.

Khan, Alisa, Mari M. Nakamura, Alan M. Zaslavsky, Jisun Jang, Jay G. Berry, Jeremy Y. Feng, Mark A. Schuster. 2015. Same-hospital readmission rates as a measure of pediatric quality of care. *JAMA Pediatrics* **169**(10) 905–912.

Krikorian, Mariam L., Amanda S. Growdon, Alyna T. Chien. 2018. Assessment of hospitalist-subspecialist agreement about who should be in charge and comparison with actual assignment practices. *Hospital Pediatrics* **8**(8) 479–485.

Langfred, Claus W., Neta A. Moye. 2004. Effects of task autonomy on performance: An extended model considering motivational, informational, and structural mechanisms. *Journal of Applied Psychology* **89**(6) 934.

Lynch, Mark F, Ken M Anson, Uday Patel. 2008. Percutaneous nephrostomy and ureteric stent insertion for acute renal deobstruction consensus based guidance. *British Journal of Medical and Surgical Urology* **1**(3) 120–125.

Miskovic, Danilo, Jake Foster, Ayman Agha, Conor P Delaney, Nader Francis, Hirotoshi Hasegawa, Alexey Karachun, Seon Hahn Kim, Wai Lun Law, John Marks, et al. 2015. Standardization of laparoscopic total mesorectal excision for rectal cancer: a structured international expert consensus. *Annals of surgery* **261**(4) 716–722.

Nieuwenhuys, Angela, Sylvia Õunpuu, Anja Van Campenhout, Tim Theologis, Josse De Cat, Jean Stout, Guy Molenaers, Tinne De Laet, Kaat Desloovere. 2016. Identification of joint patterns during gait in children with cerebral palsy: a Delphi consensus study. *Developmental Medicine & Child Neurology* **58**(3) 306–313.

Perrow, Charles. 1967. A framework for the comparative analysis of organizations. *American Sociological Review* **32**(2) 194–208.

Puranam, Phanish, Oliver Alexy, Markus Reitzig. 2014. What's 'new' about new forms of organizing? *The Academy of Management Journal* **39**(2) 162–180.

Qaseem, A., D. Kansagara, J. S. Lin, R. A. Mustafa, T. J. Wilt. 2019. The development of clinical guidelines and guidance statements by the clinical guidelines committee of the american college of physicians: Update of methods. *Annals of Internal Medicine* **170**(12) 863–870.

Rahbari, Nuh N, Phillip Knebel, Markus K Diener, Christoph Seidlmayer, Karsten Ridwelski, Hartmut Stöltzing, Christoph M Seiler. 2009. Current practice of abdominal wall closure in elective surgery–is there any consensus? *BMC Surgery* **9**(1) 1–8.

Reay, Trish, Karen Golden-Biddle, Kathy Germann. 2006. Legitimizing a new role: Small wins and microprocesses of change. *The Academy of Management Journal* **49**(5) 977–998.

Roth, Aleda, Anita L. Tucker, Sriram Venkataraman, Jon Chilingerian. 2019. Being on the productivity frontier: Identifying "triple aim performance" hospitals. *Production and Operations Management* **28**(9) 2165–2183.

Saghafian, S., W.J. Hopp, S.M.R. Iravani, Y. Cheng, D. Diermeier. 2018. Workload management in telemedical physician triage and other knowledge-based service systems. *Management Science* **64**(11) 5180–5197.

Saghafian, S., W.J. Hopp, M.P. Van Oyen, J.S. Desmond, S.L. Kronick. 2014. Complexity augmented triage: A tool for improving patient safety and operational efficiency. *Manufacturing and Service Operations Management* **16**(3) 329–345.

Saghafian, Soroush, Wallace J Hopp. 2019. The role of quality transparency in health care: Challenges and potential solutions. *National Academy of Medicine (NAM) Perspectives* **2019**.

Saghafian, Soroush, Wallace J Hopp. 2020. Can public reporting cure healthcare? the role of quality transparency in improving patient–provider alignment. *Operations Research* **68**(1) 71–92.

Saghafian, Soroush, Derya Kilinc, Stephen J. Traub. 2022a. Dynamic assignment of patients to primary and secondary inpatient units: Is patience a virtue? *Working Paper (HKS Working Paper No. RWP17-010), Harvard University, Cambridge, MA.* .

Saghafian, Soroush, Lina Song, Joseph P. Newhouse, Marry B. Landrum, John Hsu. 2022b. The impact of vertical integration on physician behavior and healthcare delivery: Evidence from gastroenterology practices. *Managment Science* .

Saghafian, Soroush, Lina Song, Ali Raja. 2022c. Towards a more efficient healthcare system: Opportunities and challenges caused by hospital closures amid the COVID-19 pandemic. *Health Care Management Science* **25** 187–190.

Schultz, K.L., D.C. Juran, J.W. Boudreau. 1999. The effects of low inventory on the development of productivity norms. *Management Science* **45**(12) 1664–1678.

Schultz, K.L., D.C. Juran, J.W. Boudreau, J.O. McClain, L.J. Thomas. 1998. Modeling and worker motivation in jit productin systems. *Management Science* **44**(12) 1595–1607.

Shafritz, Jay M., J. Steven Ott, Yong Suk Jang. 2015. *Classics of Organization Theory*. Cengage.

Shumsky, R.A., E.J. Pinker. 2003. Gatekeepers and referrals in services. *Management Science* **49**(7) 839–856.

Simon, Tamara D., Jay Berry, Chris Feudtner, Bryan L. Stone, Xiaoming Sheng, Susan L. Bratton, J. Michael Dean, Rajendu Srivastava. 2010. Children with complex chronic conditions in inpatient hospital settings in the united states. *Pediatrics* **126**(4) 647–655.

Sinha, K. K., A. H. Van de Ven. 2005. Designing work within and between organizations. *Organization Science* **16**(4) 389–408.

Song, H., A.L. Tucker, R. Graue, S. Moravick, J.J. Yang. 2020. Capacity pooling in hospitals: The hidden consequences of off-service placement. *Management Science* **66**(9) 3825—3842.

Spear, Steven J, Mark Schmidhofer. 2005. Ambiguity and workarounds as contributors to medical error. *Annals of Internal Medicine* **142**(8) 627–630.

Stevens, J. P., D. J. Nyweide, S. Maresh, L. A. Hatfield, M. D. Howell, B. E. Landon. 2017. Comparison of hospital resource use and outcomes among hospitalists, primary care physicians, and other generalists. *JAMA Internal Medicine* **177**(12) 1781—-1787.

Stock, J.H., J.H. Wright, M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* **20**(4) 518–529.

Thomas, Pete, Jan Hewitt. 2011. Managerial organization and professional autonomy: A discourse-based conceptualization. *Organization Studies* **32**(10) 1373–1393.

Thompson, James D. 1967. *Organizations in Action: Social Science Bases of Administrative Theory*. Transaction Publishers.

Thornton, Patricia H., Candance Jones, Kenneth Kury. 2005. Institutional logics and institutional change in organizations: Transformation in accounting, architecture, and publishing. *Research in the Sociology of Organizations* **23** 125–170.

Traub, S.J., C. Stwart, R. Didehban, A. Bartley, S. Saghafian. 2016. Emergency department rotational patient assignment. *Annals of Emergency Medicine* **67**(2) 206–215.

Tucker, Anita L, Sarah Zheng, John W Gardner, Roger E Bohn. 2020. When do workarounds help or hurt patient outcomes? the moderating role of operational failures. *Journal of Operations Management* **66**(1-2) 67–90.

Valeri, L., T. J. VanderWeele. 2013. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods* **18**(2) 137–150.

van de Velde, Cornelis JH, Petra G Boelens, Josep M Borras, Jan-Willem Coebergh, Andres Cervantes, Lennart Blomqvist, Regina GH Beets-Tan, Colette BM van den Broek, Gina Brown, Eric Van Cutsem, et al. 2014. EURECCA colorectal: multidisciplinary management: European consensus conference colon & rectum. *European Journal of Cancer* **50**(1) 1–e1.

van Donselaar, K., V. Gaur, T. van Woensel, R. Broekmeulen, J. Fransoo. 2010. Ordering behavior in retail stores and implications for automated replenishment. *Management Science* **56**(5) 766–784.

VanderWeele, Tyler. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

von Hippel, Eric. 1990. Task partitioning: An innovation process variable. *Research Policy* **19**(5) 407–418.

Waldman, Amy T, Mark P Gorman, Mary R Rensel, Tracy E Austin, Deborah P Hertz, Nancy L Kuntz, Network of Pediatric Multiple Sclerosis Centers of Excellence of the National Multiple Sclerosis Society. 2011. Management of pediatric central nervous system demyelinating disorders: consensus of United States neurologists. *Journal of Child Neurology* **26**(6) 675–682.

Appendix: Additional Robustness Checks Using IV Analysis

Our selected IV is whether a generalist (versus specialist) was assigned. We choose this variable as our instrument, partially because (a) it is correlated with assignment matching guideline ($corr = 0.44$, $p < 0.001$), and (b) is not correlated with one of our main outcome variables, 30-day readmissions ($corr = 0.02$, $p < 0.001$). We observed that our chosen instrument is to some extent correlated with some of the other outcome variables. Thus, we focused on carrying out the IV analysis to test the robustness of our findings with respect to 30-day readmissions, which is one of the most important metrics for most hospitals.

We conceptualize our approach using the Directed Acyclic Graph (DAG) depicted in Figure EC.1. In this figure, the treatment variable T represents when assignment matched the guideline (preferred assignment as determined by the consensus opinion), the instrumental variable Z represents when a generalist was assigned, X denotes our vector of controls, Y represents the outcome variable (30-day readmission), and e denotes the error term variable. As this figure shows, we consider the treatment variable (T) as endogenous and utilize our instrument (Z) to adjust for it. To do so, we perform 2-stage models, with the first model regressing the potentially endogenous variable T on Z and X. In the second stage, we use GLM to fit each of the outcomes on X and the fitted values of T in the first stage (which corrected for endogeneity). We also cluster each model by physician, as we did previously.

From a theoretical argument perspective, we believe our IV approach described above is valid for the following reasons. First, our outcome variable (30-day readmission) is mainly related to patient conditions (as opposed to the type of provider that is originally assigned). Patient conditions are captured in variables such as the chronic condition indicator (CCI), age, and other factor that already serve as controls in our analyses. For example, we observe from our data that the 30-day readmission rate among patients for whom a generalist is assigned is similar to those for whom a specialist is assigned (7.2% and 6.2%, $p = 0.18$). Furthermore, patients across these groups seem to be similar in terms of important indicators such as CCI (1.58 and 1.52, $p > 0.95$). Thus, when it comes to our outcome of interest, 30-day readmission, it seems that our instrument is (to a great extent) as good as random assignment. Next, we argue that there is a causal path between our IV and our outcome variable of interest that passes through our treatment variable. This is because when a deviation occurs (see the treatment variable in Figure EC.1), the patient's course of care may be altered, which is likely the reason deviating from or adhering to preferred assignment (treatment variable) can influence 30-day readmission (outcome variable). The rate of deviations, however, differs depending on whether a generalist of specialists is originally assigned, which is why our IV is relevant. Finally, it is relatively reasonable (at least based the observations above) to

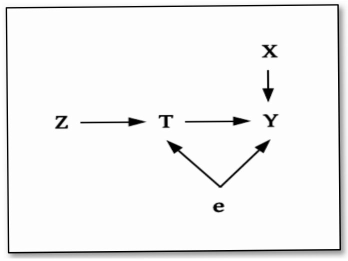


Figure EC.1 The Directed Acyclic Graph (DAG) Behind Our Instrumental Variable Approach [Z: generalist is assigned; T: assignment matches the guideline; Y: outcome variable (30-day readmissions); X: vector of control variables].

assume that the above-mentioned causal path between our outcome variable and IV (that passes through our treatment variable) is the only causal path that connects the two.

More formally, our tests (e.g., Sargan–Hansen) show that, for 30-day readmissions, our chosen instrument likely satisfies the exclusion restriction.[9] Further, when using this instrument, we find an F-statistic of 435 ($p < 0.001$) for the first-stage estimator, which surpasses the minimum threshold F-statistic of 10 typically required for identifying a weak instrument (see, e.g., Stock et al. 2002).[10] Overall, our statistical tests show that our chosen instrument is not weak, and can be reliably used as a robustness check to investigate the validity of our findings, especially with respect to the findings related to the most important outcome variable in our study (30-day readmissions).

Our results are presented in Table EC.8. As this table shows, we observe similar results to our earlier ones (see, e.g., Table 4). In particular, the implications of following the preferred assignment on performance outcomes is consistent with our earlier results. Notably, we do not observe any effect changing direction (from positive to negative or from negative to positive) when using the IV approach. This, along with the fact that our IV is not weak and strongly satisfies the exclusion restriction for our main outcome variable (30-day readmission), gives us confidence that our results are relatively robust, and not biased due to potential endogeneity issues.

However, as is often the case when using an IV, we cannot *prove* that our IV is fully valid, though all our tests and investigations indicate that it likely is. We leave it to future research to further validate our findings through randomized controlled trials or by obtaining other data sets that might allow using other variables as an IV.

[9] For some other outcome variables we study, we observe that the exclusion restriction might not hold as strongly. However, in rerunning our analyses using our IV, we noticed fairly robust results across all of the outcome variables, which gave us confidence about the validity of our main findings discussed in previous sections.

[10] We analyzed two other candidates as potential instruments: (1) whether assignment occurred between the hours of 5pm and 8am (during these times, a generalist is more likely to be assigned than a specialist), and (2) whether the patient's diagnosis at admission matched diagnosis at discharge. However, unlike our chosen IV, both of these were weak instruments and showed a low F-statistic.

Appendix: Additional Tables

Table EC.1 Performance Implication with Interaction Terms

Dependent Variable:	Operational Efficiency				Cost				Quality							
Variables	Length of stay (days) M5	Length of stay (days) M6	Length of stay (days) M7	Length of stay (days) M8	Total charges ($) M10	Total charges ($) M11	Total charges ($) M12	Total charges ($) M13	Readmission, 30-day M15	Readmission, 30-day M16	Readmission, 30-day M17	Readmission, 30-day M18	Adverse events M20	Adverse events M21	Adverse events M22	Adverse events M23
Preferred Assignment	0.313***	0.142**	0.242***	0.155***	0.320***	0.107*	0.270***	0.190***	0.348	-0.048	0.096	0.153	0.353	0.021	-0.434*	-0.060
	(0.052)	(0.051)	(0.045)	(0.034)	(0.053)	(0.053)	(0.047)	(0.035)	(0.224)	(0.216)	(0.185)	(0.166)	(0.283)	(0.220)	(0.212)	(0.179)
Type of physician assigned (specialty)	0.022	0.021	0.015	0.018	0.031*	0.026	0.023	0.025	-0.011	-0.032	-0.011	-0.010	0.157**	0.163**	0.183***	0.159**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)	(0.016)	(0.045)	(0.048)	(0.046)	(0.045)	(0.049)	(0.051)	(0.050)	(0.050)
Task categories																
Well-defined	0.145**	0.000	0.022	0.028	0.117*	-0.007	0.016	0.028	-0.565**	-0.769***	-0.764***	-0.745***	0.922***	0.628***	0.534***	0.637***
	(0.045)	(0.031)	(0.032)	(0.031)	(0.047)	(0.032)	(0.033)	(0.032)	(0.197)	(0.137)	(0.138)	(0.137)	(0.230)	(0.158)	(0.158)	(0.157)
High complexity	0.009	0.000	-0.023	-0.023	0.034	-0.031	0.008	0.007	-0.396*	-0.660**	-0.405*	-0.411*	-1.179***	-1.097***	-1.201***	-1.191***
	(0.034)	(0.047)	(0.035)	(0.034)	(0.035)	(0.049)	(0.036)	(0.035)	(0.165)	(0.240)	(0.165)	(0.164)	(0.176)	(0.238)	(0.177)	(0.175)
High contact	0.134**	0.118**	0.284***	0.094*	0.246***	0.233***	0.383***	0.204***	-0.309	-0.306	-0.472	-0.334	-0.101	-0.157	-0.931**	-0.161
	(0.043)	(0.043)	(0.065)	(0.043)	(0.044)	(0.044)	(0.068)	(0.044)	(0.198)	(0.199)	(0.281)	(0.198)	(0.209)	(0.208)	(0.309)	(0.208)
Resource-intensive	0.108*	0.102*	0.093	0.572***	0.145**	0.146**	0.130**	0.650***	-0.505*	-0.531*	-0.525*	-0.458	-0.415	-0.428	-0.445	-0.067
	(0.048)	(0.048)	(0.048)	(0.071)	(0.050)	(0.050)	(0.050)	(0.074)	(0.236)	(0.237)	(0.236)	(0.373)	(0.269)	(0.270)	(0.269)	(0.410)
Interactions																
Well-defined x Routine assignment	-0.268***				-0.222***				-0.349				-0.590			
	(0.061)				(0.064)				(0.269)				(0.314)			
High complexity x Routine assignment		0.005				0.128				0.446				-0.195		
		(0.066)				(0.068)				(0.303)				(0.325)		
High contact x Routine assignment			-0.248***				-0.225**				0.239				1.288***	
			(0.073)				(0.077)				(0.329)				(0.355)	
Resource-intensive x Routine assignment				-0.745***				-0.812***				-0.101				-0.625
				(0.087)				(0.091)				(0.434)				(0.505)
Constant	1.767***	1.908***	1.877***	1.891***	10.464***	10.633***	10.550***	10.563***	-3.244**	-2.915**	-3.061**	-3.097**	-2.948**	-2.708**	-2.538**	-2.649**
	(0.214)	(0.215)	(0.213)	(0.212)	(0.226)	(0.227)	(0.226)	(0.224)	(1.125)	(1.126)	(1.121)	(1.118)	(0.950)	(0.937)	(0.941)	(0.928)
N	4729	4729	4729	4729	4729	4729	4729	4729	4729	4729	4729	4729	4729	4729	4729	4729

Notes: Generalized linear model results reported (M1-M8: gamma family, logistic link; M9-M16: binomial family, logit link). Standard errors are in parentheses. Model is adjusted by patient characteristics and patient volume, and clustered by physician assigned. Includes year fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table EC.2 Mediation Effect

Mediation Analysis	Well- Defined Tasks (n = 3,222)				High Complexity Tasks (n = 2,289)				High Contact Tasks (n = 1,686)				Resource-Intensive Tasks (n = 1,806)			
	Operational Efficiency	Cost	Quality		Operational Efficiency	Cost	Quality		Operational Efficiency	Cost	Quality		Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse events	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse events	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse events	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse events
Preferred assignment→ Advanced imaging	-0.291** (0.079)	-0.291** (0.079)	-0.291** (0.079)	-0.291** (0.079)	-0.053 (0.104)	-0.053 (0.104)	-0.053 (0.104)	-0.053 (0.104)	-0.503*** (0.129)	-0.503*** (0.129)	-0.503*** (0.129)	-0.503*** (0.129)	-0.294* (0.120)	-0.294* (0.120)	-0.294* (0.120)	-0.294* (0.120)
Advanced imaging→ Outcome	0.996*** (0.015)	68046*** (9263)	-0.068 (0.207)	-0.194 (0.177)	0.870*** (0.017)	80236*** (12869)	-0.311 (0.260)	0.058 (0.256)	1.138*** (0.021)	106792*** (16719)	0.020 (0.297)	-0.135 (0.145)	0.938*** (0.019)	87622*** (15401)	-0.066 (0.289)	-0.151 (0.303)
MTE	1.161*** (0.032)	26257** (8146)	0.792 (0.210)	0.806 (0.202)	1.197*** (0.033)	34881* (13453)	1.262 (0.229)	0.865 (0.242)	1.275*** (0.049)	57663** (19739)	0.907 (0.303)	1.050 (0.008)	0.980 (0.045)	19608 (17136)	1.151 (0.289)	1.021 (0.303)
NDE	1.235*** (0.017)	29650*** (7456)	0.789 (0.208)	0.799 (0.200)	1.210*** (0.018)	35694* (11962)	1.258 (0.227)	0.865 (0.239)	1.458*** (0.023)	68417*** (17255)	0.909 (0.302)	1.052 (0.008)	1.046 (0.021)	24931 (15757)	1.147 (0.287)	1.012 (0.299)
NIE	0.940* (0.027)	-3393** (1271)	1.003 (0.903)	1.009 (0.026)	0.990 (0.028)	-813 (1608)	1.003 (0.027)	0.998 (0.028)	0.875** (0.044)	-10754** (3283)	0.998 (0.052)	0.998 (0.018)	0.937 (0.039)	-5323* (2397)	1.004 (0.042)	1.009 (0.042)
Bias-corrected 95% confidence interval (NIE)	[0.892, 0.990]	[-5833, -903]	[0.951, 1.058]	[0.958, 1.063]	[0.938, 1.045]	[-3964, 2339]	[0.950, 1.058]	[0.947, 1.055]	[0.802, 0.953]	[-17190, -4320]	[0.901, 1.106]	[0.985, 1.002]	[0.868, 1.012]	[-10022, -624]	[0.924, 1.091]	[0.929, 1.096]
Mediating effect	YES	YES	No	No	No	No	No	No	YES	YES	No	No	No	YES	No	No

Notes: MTE = marginal total effect; NDE = natural direct effect; NIE = natural indirect effect; Standard errors are in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table EC.3 Mediation Effect Heterogeneity

Mediation Analysis	Generalist Assigned (n = 1,477)				Specialist Assigned (n = 3,252)			
	Operational Efficiency	Cost	Quality		Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse events	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse events
Preferred assignment → Advanced imaging	-1.502*** (0.232)	-1.502*** (0.232)	-1.502 (0.232)	-1.502*** (0.232)	0.544*** (0.103)	0.544*** (0.103)	0.544*** (0.103)	0.544*** (0.103)
Advanced imaging → Outcome	1.007*** (0.025)	53474*** (4663)	0.329 (0.286)	0.259 (0.316)	0.992*** (0.014)	81019*** (9543)	-0.511** (0.196)	-0.245 (0.166)
Total effect	0.534*** (0.123)	-31400** (10046)	1.431 (0.551)	0.439 (0.519)	1.561*** (0.025)	45514*** (12173)	1.103 (0.203)	2.110*** (0.194)
NDE	0.757*** (0.043)	-14577* (7307)	1.298 (0.544)	0.406 (0.509)	1.382*** (0.017)	37199*** (10543)	1.155 (0.203)	2.160*** (0.194)
NIE	0.705** (0.116)	-16824*** (3267)	1.102 (0.138)	1.081 (0.145)	1.129*** (0.019)	8315*** (1882)	0.955 (0.027)	0.977 (0.027)
Bias-corrected 95% confidence interval (NIE)	[0.615, 0.798]	[-26951, -10162]	[0.948, 1.329]	[0.870, 1.295]	[1.075, 1.192]	[4704 - 13662]	[0.914, 0.986]	[0.940, 1.005]
Mediating effect	YES	YES	No	No	YES	YES	No	No

Notes: NDE = natural direct effect; NIE = natural indirect effect; Standard errors are in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table EC.4 Balance of Covariates

Patient characteristics	Standard difference	Variance ratio
Patient's age	-0.000	0.999
Male patient	-0.004	1.000
Private insurance	0.003	1.002
Chronic condition indicator (CCI)	0.001	1.003

Table EC.5 Summary of Hypothesis Tests, using 45% Agreement Level to Determine Preferred Assignment

Hypothesis #	Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission rate, 30-day	Adverse event
1	F	–	–	F
2	T	T	T	F
3	T	T	–	T
4	T	T	–	F

Notes: T = true; F = false; – = insignificant results.

Table EC.6 Summary of Hypothesis Tests, Using 55% Agreement Level to Determine Preferred Assignment

Hypothesis #	Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission rate, 30-day	Adverse event
1	–	–	T	–
2	F	F	–	F
3	–	–	–	–
4	T	T	–	T

Notes: T = true; F = false; – = insignificant results.

Table EC.7 Summary of Hypothesis Tests, Using 0.5 Reliability Coefficient Agreement Level to Determine Preferred Assignment

Hypothesis #	Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission rate, 30-day	Adverse event
1	F	F	–	T
2	–	–	–	F
3	T	T	–	T
4	T	T	–	–

Notes: T = true; F = false; – = insignificant results.

Table EC.8 Summary of Performance Implications Across Task Categories, Using an Instrumental Variable Approach

Dependent Variable:	Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse event
Well-defined task	-	-	-	+
High complexity task	+	+	-	-
High contact task	0	+	-	0
Resource-intensive task	0	+	-	0

Notes: + = positive effect; - = negative effect; 0 = no effect.

Table EC.9 List of the Diagnoses in the Survey

No.	Description	Division
1	ADEM, encephalopathy	Neurology
2	Afebrile (unprovoked seizures)	Neurology
3	Brain mass, benign	Neurology
4	Guillain-Barre, CIDP, other demyelinating disorders	Neurology
5	Infantile botulism	Neurology
6	Infantile spasms	Neurology
7	Movement disorders, dystonia	Neurology
8	Myasthenia gravis	Neurology
9	Neuropathy/neuritis (mono or poly) - acute	Neurology
10	Neuropathy/neuritis (mono or poly) - chronic	Neurology
11	Opsoclonus-myoclonus	Neurology
12	Seizure - admit for video EEG	Neurology
13	Seizure/epilepsy - known for breakthrough or intractable, vs chronic epilepsy in patient admitted for other diagnosis (i.e. pneumonia)	Neurology
14	Stroke - known, with related problem or treatment	Neurology
15	Stroke - new work-up, or with unrelated problem	Neurology
16	Weakness - acute	Neurology
17	Weakness and/or hypotonia - chronic	Neurology
18	Abnormal thyroid function test	Endocrinology
19	Adrenal insufficiency	Endocrinology
20	Ambiguous genitalia	Endocrinology
21	Atrophy of testis	Endocrinology
22	Congenital adrenal hyperplasia	Endocrinology
23	Congenital anomaly of adrenal glandCongenital anomaly of adrenal gland	Endocrinology
24	Craniopharyngioma	Endocrinology
25	Cushing syndrome	Endocrinology
26	Cyst of thyroid	Endocrinology
27	Diabetes insipidus	Endocrinology
28	Diabetes insipidus, nephrogenic	Endocrinology
29	Diabetes mellitus - general pediatric problem	Endocrinology
30	Diabetes mellitus - problem relating to underlying disease	Endocrinology
31	Diabetic ketoacidosis	Endocrinology
32	Disorder of male genital organ	Endocrinology
33	Disorder of menstruation & abnormal bleeding from female genital tract	Endocrinology
34	Electrolyte disturbance related to an endocrine disorder	Endocrinology
35	Galactorrhea	Endocrinology
36	Hyperaldosteronism	Endocrinology
37	Hypercalcemia	Endocrinology
38	Hyperparathyroidism	Endocrinology
39	Hyperthyroidism	Endocrinology
40	Hypoaldosteronism	Endocrinology
41	Hypocalcemia	Endocrinology
42	Hypoparathyroidism	Endocrinology
43	Hypothyroidism	Endocrinology
44	Panhypopituitarism	Endocrinology
45	Secondary diabetes	Endocrinology
46	Thyroid cancer	Endocrinology
47	Thyroid storm	Endocrinology
48	Toxic diffuse goiter, without thyrotoxic crisis/storm	Endocrinology
49	Unspecified congenital anomaly of genital organs	Endocrinology
50	Unspecified disorder of adrenal gland	Endocrinology
51	Acute liver failure	Gastroenterology
52	Ascites	Gastroenterology
53	Attention to colostomy, ileostomy, J-tube, G-tube	Gastroenterology
54	Cholangitis	Gastroenterology
55	Chronic liver disease	Gastroenterology
56	Complication due to CVC	Gastroenterology

57	Congenital biliary atresia	Gastroenterology
58	Conjugated hyperbilirubinemia	Gastroenterology
59	Constipation, NOS	Gastroenterology
60	Constipation, related to chronic GI disease	Gastroenterology
61	Crohn's disease - presenting with complication of disease	Gastroenterology
62	Crohn's disease, stable - presenting with general pediatric problem	Gastroenterology
63	Dysphagia	Gastroenterology
64	Esophagitis	Gastroenterology
65	Gastroesophageal reflux	Gastroenterology
66	Hematemesis	Gastroenterology
67	Hematochezia	Gastroenterology
68	Hepatitis, acute	Gastroenterology
69	Hepatitis, autoimmune	Gastroenterology
70	Intestinal dysmotility - small bowel	Gastroenterology
71	Intestinal dysmotility - colon	Gastroenterology
72	Intestinal malabsorption	Gastroenterology
73	Intestinal pseudo-obstruction	Gastroenterology
74	Pancreatitis, acute	Gastroenterology
75	Pancreatitis, chronic	Gastroenterology
76	Persistent vomiting	Gastroenterology
77	Portal hypertension	Gastroenterology
78	Short Bowel Syndrome - presenting with complications of disease	Gastroenterology
79	Short Bowel Syndrome, stable - presenting with general pediatric problem or complication due to CVC	Gastroenterology
80	Ulcerative colitis - presenting with complication of disease	Gastroenterology
81	Ulcerative colitis, stable - presenting with general pediatric problem	Gastroenterology
82	Anemia - new workup	Hematology-Oncology
83	Anemia - known d/o, admitted with heme issue	Hematology-Oncology
84	Anemia - known d/o admitted with gen peds issue	Hematology-Oncology
85	Anemia due to any reason, that is severe and requiring blood products	Hematology-Oncology
86	Neutropenia - new workup	Hematology-Oncology
87	Neutropenia - known d/o admitted with heme issue	Hematology-Oncology
88	Neutropenia - known d/o admitted with gen peds issue	Hematology-Oncology
89	Thrombocytopenia - new workup	Hematology-Oncology
90	Thrombocytopenia - known d/o admitted with heme issue	Hematology-Oncology
91	Thrombocytopenia - known d/o admitted with gen peds issue	Hematology-Oncology
92	Combo of cytopenias - with HSM, mediastinal mass, lymphadenopathy, or peripheral blasts - new workup	Hematology-Oncology
93	Combo of cytopenias - no HSM, no lymphadenopathy, no peripheral blasts - new workup	Hematology-Oncology
94	Combo of cytopenias - known d/o, admitted with heme related issue	Hematology-Oncology
95	Combo of cytopenias - known d/o, admitted with gen peds issue	Hematology-Oncology
96	Factor disorders, admitted for any reason	Hematology-Oncology
97	SS, SC, or Sthal admitted for any reason	Hematology-Oncology

98	Thrombotic disorders, admitted for gen peds issue	Hematology-Oncology
99	Thrombotic disorders, new diagnosis admitted for work-up or known dx admitted for heme issue	Hematology-Oncology
100	Histiocytic disorders (HLH), admitted for any reason	Hematology-Oncology
101	Histiocytic disorders (LCH), admitted for any reason	Hematology-Oncology
102	Bloodstream infection - known hematologic/onc. disorder and CVC	Hematology-Oncology
103	Active malignancy, on treatment or end stage, admitted with any issue, including but not limited to, chemotherapy, fever and neutropenia - but excludes rehab	Hematology-Oncology
104	Active malignancy, admitted for rehab	Hematology-Oncology
105	History of malignancy, now in remission and not getting therapy, less than one year off therapy with gen peds issue	Hematology-Oncology
106	History of malignancy, now in remission and not getting therapy, more than one year off therapy with gen peds issue	Hematology-Oncology
107	History of malignancy, now in remission and not getting therapy, more than one year off therapy, admitted with original presenting symptoms (relapse possibility)	Hematology-Oncology
108	History of malignancy, now in remission and not getting therapy, more than one year off therapy, admitted with known side effects of prior chemotherapy	Hematology-Oncology
109	Lymphadenopathy without mediastinal mass or cytopenias or peripheral blasts	Hematology-Oncology
110	Lymphadenopathy with mediastinal mass or cytopenias or peripheral blasts	Hematology-Oncology
111	New mass work-up	Hematology-Oncology
112	New intracranial mass, no diagnosis established	Hematology-Oncology
113	New intracranial mass, incompletely resected (benign)	Hematology-Oncology
114	New intracranial mass, incompletely resected (malignant)	Hematology-Oncology
115	Chronic lung disease patient, followed by the pulm service, admitted for exacerbation (includes RAD, pneumonia, etc.)	Pulmonology
116	Chronic lung disease patient, not followed by the pulm service, admitted for exacerbation (includes RAD, pneumonia, etc.)	Pulmonology
117	Chronic lung disease of infancy/BPD preparing for first discharge	Pulmonology
118	Chronic Lung Disease of infancy/BPD with intercurrent illness	Pulmonology
119	Cystic fibrosis with admission not related to underlying CF (e.g. cellulitis)	Pulmonology
120	Cystic fibrosis with admission related to underlying CF (e.g. pneumonia, pancreatitis)	Pulmonology
121	New home ventilator discharge	Pulmonology
122	Obstructive sleep apnea, newly diagnosed admitted for management	Pulmonology
123	Obstructive sleep apnea, previously diagnosed admitted for further management	Pulmonology
124	Pt's with assisted ventilation (via trach/mask, CPAP, BPaP, diaphragm pacing) admitted with gen peds issue	Pulmonology

125	Pt's with assisted ventilation (via trach/mask, CPAP, BPaP, diaphragm pacing) admitted with issue related to the underlying pulm disease	Pulmonology
126	Pulmonary hypoplasia/congenital diaphragmatic hernia/other pulm anatomical abnormality	Pulmonology
127	Tracheitis with ventilator dependence	Pulmonology
128	Tracheitis without ventilator dependence	Pulmonology
129	Ankylosing spondylitis	Rheumatology
130	Arteritis	Rheumatology
131	Behcets	Rheumatology
132	Chronic nonbacterial osteomyelitis (CNO)	Rheumatology
133	Chronic pain	Rheumatology
134	Circumscribed scleroderma	Rheumatology
135	Dermatomyositis	Rheumatology
136	Familial mediterranean fever	Rheumatology
137	Fibromyalgia	Rheumatology
138	Overlap syndrome	Rheumatology
139	Pauciarticular jia	Rheumatology
140	Polyarteritis nodosa	Rheumatology
141	Polyarticular jia	Rheumatology
142	Psoriatic arthropathy	Rheumatology
143	Rheumatoid arthritis (RA)	Rheumatology
144	Spondylopathy	Rheumatology
145	Systemic jia	Rheumatology
146	Systemic lupus erythematosus	Rheumatology
147	Systemic sclerosis	Rheumatology
148	Takayasus	Rheumatology
149	Uveitis acute	Rheumatology
150	Uveitis chronic	Rheumatology
151	Wegeners granulomatosis	Rheumatology
152	Arrhythmia - new	Cardiology
153	Mechanical complication due to cardiac pacemaker	Cardiology
154	Kawasaki's Disease	Cardiology
155	Heart Transplant	Cardiology
156	Heart Murmur	Cardiology
157	Endocarditis	Cardiology
158	Disorder cardiac valve(s) - presenting with general pediatric problem	Cardiology
159	Disorder cardiac valve(s) - presenting with complication of disease	Cardiology
160	Disorder cardiac valve(s) - new diagnosis	Cardiology
161	Cyanosis	Cardiology
162	Congestive heart failure	Cardiology
163	Congenital Heart Disease ‰ÛÒ new	Cardiology
164	Congenital heart disease - presenting with general pediatric problem	Cardiology
165	Congenital heart disease - presenting with complication of underlying disease	Cardiology
166	Congenital heart block	Cardiology
167	Congenital coronary artery anomaly - presenting with general pediatric problem	Cardiology
168	Complication due to heart valve prosthesis	Cardiology
169	Chest Pain	Cardiology
170	Cardiomyopathy	Cardiology
171	Arrhythmia - stable, presenting with general pediatric problem	Cardiology
172	Myocarditis	Cardiology
173	Pericardial effusion	Cardiology
174	S/P Cardiac Catheterization	Cardiology
175	S/P Cardiac Surgery	Cardiology
176	Syncope	Cardiology

Table EC.10 Full Version of Table 2

	Preferred Assignment
Type of physician assigned (specialty)	-0.201***
	(0.0496)
Well-defined task	-0.126
	(0.100)
High complexity task	1.071***
	(0.116)
High contact task	-0.524***
	(0.140)
Resource-intensive task	0.209
	(0.154)
Patient volume	-0.0293
	(0.0300)
Patient age (years)	-0.0475***
	(0.00754)
Male patient	-0.139
	(0.0904)
Private Insurance	0.00436
	(0.0977)
Chronic condition indicator (CCI) (base CCI = 0)	
1	0.381**
	(0.144)
2	0.358*
	(0.150)
3	0.0634
	(0.175)
4	-0.0926
	(0.234)
Year (base year = 2009)	
2010	0.0692
	(0.195)
2011	0.199
	(0.190)
2012	0.315
	(0.188)
2013	0.235
	(0.188)
2014	0.257
	(0.190)
2015	0.135
	(0.204)
Flu season	-0.0198
	(0.0921)
Weekend discharge	-0.128
	(0.113)
Shift change	-0.246
	(0.338)
Discharge time stamp (base time stamp = 12 - 8am)	
8am - 1pm	-0.210
	(0.609)
1 - 4pm	-0.208
	(0.606)
4 - 7pm	-0.136
	(0.606)
7pm - 12am	-0.388
	(0.613)
Constant	0.977
	(0.674)
N	4729

Table EC.11 Full Version of Table 3

	Length of stay (days)	Total charges ($)	Readmission, 30-day	Adverse event
Preferred Assignment	0.144***	0.181***	0.155	-0.0559
	(0.0344)	(0.0355)	(0.166)	(0.180)
Type of physician assigned (specialty)	0.0210	0.0293	-0.0105	0.156**
	(0.0140)	(0.0155)	(0.0451)	(0.0496)
Well-defined task	0.000253	-0.00254	-0.749***	0.618***
	(0.0311)	(0.0321)	(0.136)	(0.157)
High complexity task	0.00254	0.0314	-0.411*	-1.195***
	(0.0345)	(0.0356)	(0.164)	(0.175)
High contact task	0.118**	0.233***	-0.330	-0.146
	(0.0431)	(0.0444)	(0.197)	(0.208)
Resource-intensive task	0.102*	0.139**	-0.525*	-0.439
	(0.0480)	(0.0496)	(0.236)	(0.270)
Patient volume	0.00943	0.0142	-0.0195	0.0343
	(0.00873)	(0.00901)	(0.0409)	(0.0434)
Patient age (years)	-0.00282	0.00640**	0.0569***	0.00651
	(0.00223)	(0.00231)	(0.00987)	(0.0107)
Male patient	-0.0392	-0.0158	-0.809***	-0.0641
	(0.0265)	(0.0272)	(0.138)	(0.132)
Private Insurance	-0.0943***	-0.0661*	0.430***	0.0302
	(0.0285)	(0.0293)	(0.131)	(0.140)
Chronic condition indicator (CCI) (base CCI = 0)				
1	0.312***	0.412***	-0.305	-0.118
	(0.0422)	(0.0434)	(0.198)	(0.199)
2	0.526***	0.601***	0.0815	0.288
	(0.0447)	(0.0459)	(0.199)	(0.202)
3	0.588***	0.686***	0.0114	0.0218
	(0.0525)	(0.0540)	(0.236)	(0.249)
4	0.526***	0.542***	0.345	-0.928
	(0.0712)	(0.0729)	(0.292)	(0.477)
Year (base year = 2009)				
2010	-0.0641	-0.0662	-0.267	-0.443
	(0.0560)	(0.0577)	(0.342)	(0.270)
2011	-0.0735	-0.0712	0.411	-0.321
	(0.0541)	(0.0557)	(0.289)	(0.256)
2012	-0.152**	-0.0321	0.475	-0.348
	(0.0533)	(0.0550)	(0.285)	(0.253)
2013	-0.154**	0.0723	0.521	-0.367
	(0.0542)	(0.0558)	(0.284)	(0.257)
2014	-0.300***	0.0126	0.341	-0.184
	(0.0550)	(0.0569)	(0.288)	(0.251)
2015	-0.307***	0.0623	0.654*	-0.364
	(0.0598)	(0.0618)	(0.299)	(0.284)
Flu season	-0.00874	-0.0177	-0.0355	0.0921
	(0.0266)	(0.0273)	(0.129)	(0.133)
Weekend discharge	-0.186***	-0.179***	-0.00909	-0.268
	(0.0327)	(0.0336)	(0.156)	(0.168)
Shift change	-0.116	-0.151	0.357	-0.158
	(0.0981)	(0.101)	(0.440)	(0.529)
Discharge time stamp (base time stamp = 12 - 8am)				
8am - 1pm	-0.558**	-0.600**	0.369	-0.505
	(0.198)	(0.211)	(1.065)	(0.868)
1 - 4pm	-0.530**	-0.540*	0.270	-0.459
	(0.197)	(0.210)	(1.062)	(0.863)
4 - 7pm	-0.579**	-0.593**	0.375	-0.424
	(0.197)	(0.210)	(1.062)	(0.864)
7pm - 12am	-0.653**	-0.659**	0.421	-0.705
	(0.199)	(0.211)	(1.068)	(0.876)
Constant	1.906***	10.58***	-3.096**	-2.648**
	(0.213)	(0.225)	(1.119)	(0.930)
N	4729	4729	4729	4729