

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

Mariam K. Atkinson¹, Soroush Saghafian²

¹Department of Health Policy and Management, T.H. Chan School of Public Health, Harvard University, Boston, MA 02115

²Harvard Kennedy School, Harvard University, Cambridge, MA 02138

Traditional principles in operations management and organizational theory suggest that standardizing assignments of tasks to individuals can have substantial benefits and boost performance. However, in various organizations including hospitals, individuals are not forced to follow recommended assignments, and thus, deviations from routine task assignments are often observed. This is due to the conventional wisdom that professionals should be given the flexibility to deviate from routine assignments when they perceive it to be advantageous. 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 by enforcing routine assignments. To perform our analyses, we identify 73 top medical diagnoses and use detailed patient-level electronic medical record (EMR) data of more than 4,700 hospitalizations. In parallel, we conduct a carefully-designed survey of physicians and utilize it to identify the routine provider type that *should have been* assigned to each patient. Using these two sources of data, we examine the consequence of deviations from routine provider assignments on three sets of performance measures: operational efficiency (measured by length of stay), quality of care (measured by the occurrence of a 30-day readmission and adverse events), and cost (measured by total charges). Taken together, our findings suggest that allowing providers to deviate from routine 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 find that routines should be enforced: deviations are either detrimental or yield no tangible benefits. To understand the 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 theory: 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 and assist them in deciding when and how to enforce routine assignments, we consider counterfactual scenarios corresponding to imposing the routine assignments fully or partially, and perform cost-effectiveness analyses. Our results indicate that enforcing routine 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; Routines; Patient-Provider Assignment; Division of Labor

History: Version: August 13, 2020.

1. Introduction

Motivation. In the professional work context, routines are repetitive or recognizable patterns of independent actions that often are the basis of protocols, procedural guidelines, and practice standards (Greenwood et al. 2019, Woolf 1992, Cyert and March 1963). Routines are an important source of efficiency and reliability in organizational performance (Feldman and Pentland 2003, Nelson and Winter 1982). Recognizing this, organizations implement coordination mechanisms to formalize processes and move toward more routine practices, which can aid in reducing unnecessary variability and improve performance. Through such coordination mechanisms, the role of routine versus deviating practices can have important bearing on performance outcomes. For example, as professionals gain more experience, they tend to deviate from routine practices in batching and sequencing tasks for completion, which ultimately diminishes productivity (Ibanez et al. 2017). In another example, the loss of practice routines due to the transfer of a central manager to a competing organization diminishes performance (Briscoe and Rogan 2015, Aime et al. 2010).

In contexts like professional organizations, routines frequently emerge through task assignments, specifically in the division of expert knowledge and invoking roles relevant to a given context. While deviations from routine task assignment can have advantages and drawbacks (Becker 2004, Timmermans and Angell 2001), the ability of professionals to balance both forms of decision-making is essential to professionalization and performance (Spee et al. 2016, Feldman and Rafaeli 2002). To reduce role ambiguity and establish measures for accountability, professionals attempt to induce consistency through routines guiding who should be in charge of a particular task (Stinchcombe 2001, Kahn et al. 1964, Borgatta and Bales 1953). Yet at times, professional expertise often involves invoking appropriate discretion to deviate from routine task assignment so that performance does not suffer (Fong Boh et al. 2007).

Realizing this tension, some theories suggest an “optimal” level of upholding routines and standards of practice in the professional context (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 establishing routines and standardizing how tasks are routed to different servers have important advantages (for studies on optimal routing in service systems with heterogeneous servers, see, e.g., Armony and Ward 2010, 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 the operations of hospitals (see, e.g., Traub et al. 2016). Due to these con-

flicting arguments, it is unclear whether and when enforcing routines (i.e., avoiding any deviation from a predefined standard practice) is beneficial.

We are particularly motivated by the assignment of generalists and specialists in our partner hospital, which is a children’s hospital on the west coast of the United States. The hospital administrators would like to know whether developing and enforcing guidelines on the assignment of generalist versus specialist physicians (for each arriving patient with certain medical conditions) based on routine practice as represented by consensus opinion of medical experts would be beneficial. Our study allows our partner hospital, in addition to many other hospitals dealing with similar issues, to gain an understanding of the impact of developing and enforcing predetermined routine assignments. Specifically, the central question we aim to answer is: *What is the impact of following routine assignments (identified based on the consensus opinion of medical experts) on various performance outcomes? Conversely, when is deviation from such routine assignments beneficial?*

Data and Research Setting. 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.¹ Separately, 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 such patients. Consistent with prior research (see, e.g., Black et al. 1999, Murphy et al. 1998), we made use of a few approaches in evaluating consensus opinion, and developed a variable for routine assignment for each medical diagnosis that represented the generally accepted opinion about generalist versus specialist assignment based on our survey results. We compared actual generalist and specialist assignments as documented in the medical records to routine assignment as indicated by the survey responses, and analyzed resulting performance outcomes along three dimensions: *operational efficiency* (measured by length of stay²), *quality* (measured by the occurrence of 30-day readmissions adverse events), and *cost* (measured by total charges). We also assessed outcomes based on the interaction between routine assignment by diagnosis (i.e., task) characteristics, focusing on the following types of tasks: *well-defined*, *high complexity*, *high contact*, and *resource-intensive*. Finally, we performed mediation analysis to understand the mechanism behind some of our results, employed cost-effectiveness analysis on some implementable counterfactual policies to provide clear recommendations for hospital administrators on when and how to enforce routine assignments, and conducted various robustness checks, including use of an instrumental variable (IV) approach and 1-nearest neighbor propensity

¹ Thus, our sample reflects a subset of all hospitalizations during this time period.

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

score matching to examine the robustness of our findings to some of our main assumptions and model specifications.

Main Findings. Our analysis of the data suggests that, for certain types of tasks, deviations³ from routine 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 associated with higher costs. Further, when tasks involve high contact, deviations from routine 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 routine assignments is beneficial for some (but not all) tasks. In particular, we find that deviating from routine assignments is beneficial 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 formally enforce guidelines reflecting routine assignments: deviations are either detrimental or come with no tangible benefits.

To shed light on a potential 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 routines. We do so by considering counterfactual scenarios corresponding to imposing the routine assignments (developed based on the consensus opinion of medical experts in our survey) fully or partially, and by performing cost-effectiveness analyses. Our results indicate that enforcing routine assignments either for all tasks or only for resource-intensive tasks is cost-effective, and that the latter—enforcing routine assignments only for resource-intensive tasks—is the superior policy.

Main Contributions. Inefficiencies with the division of labor and task assignment have been long-standing and costly challenges for managers. As organizations become more complex and new roles are introduced, causing existing roles to shift, effective task assignment only becomes more critical. We shed light on how routine versus deviating task assignment can influence outcomes, especially in relation to task type. Thus, our research provides evidence on how organizations

³ For simplicity, and based on our research goals, we use the term “deviation” throughout whenever a deviation from the consensus opinion of medical experts in our survey has occurred. However, it should be noted that the judgment call made by the decision-maker might have been considered a routine in his/her mind based on the information available to him/her at the time.

can utilize a *task-type view* to standardize assignment decisions by understanding the impact of following routines that are represented by expert opinion. In particular, our study provides a novel perspective for professional role coordination and allocation of tasks by illuminating task types for which deviations from routine assignments can be beneficial. In doing so our study offers new insights that have both managerial and theoretical implications, which we discuss next.

Managerial Implications. We provide evidence that enforcing guidelines around routine assignments can have advantages and disadvantages. Importantly, understanding the underlying tradeoffs and knowing when to formally enforce such routines 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 clear goal of improving outcomes (e.g., operational efficiency, cost, and quality) 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]

In addition, our results help hospital administrators as well as physicians to gain a better understanding of the environmental conditions under which deviations (compared to guidelines developed based on the consensus opinion of medical experts) occur more. In particular, our findings indicate that deviations are more frequent 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. Finally, our cost-effectiveness analysis show that imposing formal guidelines surrounding routine assignments either uniformly for all task types or only for resource-intensive tasks should be viewed as attractive policies by hospital administrators, since both of these policies provide an acceptable balance between patients’ health gains and the total charges accrued. Among these two, however, our results suggest that imposing such guidelines only for resource-intensive tasks will allow the hospital administrators to achieve the best outcomes.

Theoretical Implications. Our results 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 routine 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* in that 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 routine assignment 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 work 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 routine assignments. For example, our data shows that specialists deviate from routine assignments more than generalists. A reason for this could lie in the way professional status manifests in the workplace. Professional power is characteristic of roles with greater specialization, and hence, resides largely among specialist professionals (Waring and Currie 2009, Freidson 1970). Specialists' ability to use their relative status over generalist professionals can reflect in instances when specialists deflect work to generalists, thereby deviating from standardized assignment. We leave it to future research to study this or other potential reasons for deviations.

Robustness Checks. To make sure our findings are robust and are not affected by endogeneity issues, how consensus opinion is measured, the set of controls involved, or other model specifications, we perform various robustness checks. In particular, in Section 5, we re-run our analyses by making use of an instrumental variable (IV) approach, 1-nearest neighbor propensity score matching, changing the measurement for the consensus opinion in our survey including utilizing Fleiss' kappa tests as well as varying the 50% threshold used to define majority, adding other control variables to our models, and performing Bonferroni corrections to adjust for the potential dependency between our performance measures. Overall, the results of these robustness checks show that our main findings hold, which give us further confidence that are our results are fairly robust.

Organization. The rest of the paper is organized as follows. In Section 2, we describe the theoretical background of our work and state our hypotheses. In Section, 3, we discuss our research setting, data sets, and analyses. In Section 4, we present our results. Finally, we discuss our robustness checks in Section 5, elaborate on the limitations of our work in Section 6, and briefly conclude in Section 7.

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 routine client demands (Chase and Tansik 1983). Additionally, it is known that generalists have the know-how to employ a greater variety of resources to complete a task as well as deal with large quantities of information that span a variety of disciplines (Treem 2012), while specialists have the capability to troubleshoot nuanced or ambiguous issues in their knowledge domain (Boone et al. 2000).

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 task assignment 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 routine practices to enhance service and accommodate circumstances in the work environment (Reay et al. 2006). Due to the sometimes unnecessary complications of task assignment determinations, and the subsequent need to streamline professional work, organizations have attempted to routinize decisions around task assignment. Yet, professionals may still maintain their capacity for discretionary decision-making in their work. Since professionals engage in routine and deviating practices to manage task assignment (Krikorian et al. 2018), we aim to study whether and when deviation from routine assignments is beneficial.

In closing this section, we note that while little is known about when deviations from routine task assignments can yield performance improvements, deviations have been studied from other aspects, 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), and following system generated recommendations (see, e.g., van Donselaar et al. 2010). 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. Finally, 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 from routine task assignments on various dimensions of performance, including quality of care, operational efficiency, and costs.

2.2. Routine Task Assignment and Task Types

Routines can be dictated by task characteristics. Prior research describes how routines around coordination can improve performance, especially by enhancing the shared understanding about a task (Van de Ven et al. 1976). However, reliance on routines can have unintended consequences in potentially diminishing communication and coordination needed for a particular task (Orasanu 2001, Entin and Serfaty 1989, Orasanu 1993). The effect of routine practices on performance is driven by the type of task and who is enacting the practice (Hong et al. 2019, Kuntz et al. 2016). 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 the definition of 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 routine 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. First, 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 routine task assignment, meaning that following the routine assignment (rather than deviating from it) is likely to be advantageous to performance for well-defined tasks. Thus:

HYPOTHESIS 1. *For well-defined tasks, following the routine assignment 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 or routinized. 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 routine assignments. Specifically, if in deciding the appropriate provider type that should be assigned, professionals are forced to follow what the routine assignment suggests, their assignment decisions might worsen, since they may miss the opportunity to take into account the nuances of complex tasks. This leads us the following hypothesis:

HYPOTHESIS 2. *For high complexity tasks, following the routine assignment 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 routines and 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 routines, and enforcing them through disallowing potential deviations can have negative consequences. Thus, we hypothesize that:

HYPOTHESIS 3. *For high contact tasks, following the routine assignment 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 routine 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 routine assignment improves performance outcomes.*

2.3. A Potential Mechanism

Our research also explores a possible causal mechanism that could explain the relationship between routine (versus deviating) assignment and performance outcomes. Specifically, we identify a potential mediating effect that could explain (a) why when professionals in our context deviate from routine assignments, the performance outcomes are impacted, 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 routine 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 blur 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 routine assignments and removing such deviations is beneficial.

In practice, assigning tasks to generalists and specialists requires the ability to manage the tension between breadth of generalist expertise and the depth of specialist expertise based on the situation. Managing this tension involves making the decision to invoke generalist and specialist expertise with awareness of the financial and human resource implications (e.g., specialists are fewer in number and cost more to utilize than generalists). 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 electronic medical records (EMRs).

Survey of Physicians. To identify routine 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 on three hospitalists. The survey listed the “top diagnoses” for each specialty (a total of 176 diagnoses across seven specialty areas; see Table EC.7 in the Appendix for mote 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 particular 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

involving 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 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. 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. In order 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)⁴. 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 (i.e., routinized assignment) did not exist for several conditions. The EMR data included 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

⁴The 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.

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 routine (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 measured 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. 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 inputted 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. *Routine 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, routine 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.⁵ If the actual assignment recorded in the medical records deviated from that, we noted that as deviating assignment (variable takes value of zero). For the 73 diagnoses in the survey, physicians expressed that specialists should be assigned to 44 (60%) diagnoses and generalists to 29 (40%) diagnoses (based on the modal responses in the survey).

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 demonstrate high fundamentality, because they 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. 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., physi-

⁵ In our robustness checks, we alter this majority rule approach for measuring the consensus opinion, and make use of different definitions for identifying routine assignment, including utilizing Fleiss’ kappa tests as well as varying the 50% threshold used to define majority (see Section 5.3).

cal/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 routine assignment on performance depends on the type of task category, we included interactions between each of the four task categories and routine 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 gender 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 a *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). 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 controlled for 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. Finally, to control for the type of professional, we included a variable termed *generalist assigned*, which takes a value of one when a generalist is assigned and zero if a specialist is assigned.

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

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 Routine assignment	0.548	0.498	1															
2 Generalist assigned	0.312	0.463	0.4548*	1														
3 Length of stay (days)	5.900	14.774	0.0273	-0.0458*	1													
4 Total charges (\$)	54.800	200.798	0.0313*	-0.0576*	0.9317*	1												
5 Readmission, 30-day	0.065	0.246	-0.0148	0.0193	0.0334*	0.0310*	1											
6 Adverse event	0.070	0.256	-0.0754*	-0.0624*	0.0491*	0.0538*	-0.0287*	1										
7 Well-defined task	0.681	0.466	-0.0143	-0.1541*	-0.0451*	-0.0308*	-0.1060*	0.0410*	1									
8 High complexity task	0.484	0.500	0.1788*	-0.2337*	0.0346*	0.0466*	-0.0879*	-0.1343*	0.2510*	1								
9 High contact task	0.357	0.479	0.1300*	-0.2063*	0.0064	0.0390*	-0.0791*	-0.0375*	0.1168*	0.3286*	1							
10 Resource-intensive task	0.382	0.486	0.1792*	-0.2630*	0.0262	0.0449*	-0.0900*	-0.1177*	0.1630*	0.7488*	0.5144*	1						
11 Patient volume	2.123	1.570	0.0157	0.0272	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*	-0.1747*	-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	0.0214	-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*	-0.0674*	-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	0.0252	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	-0.0035	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

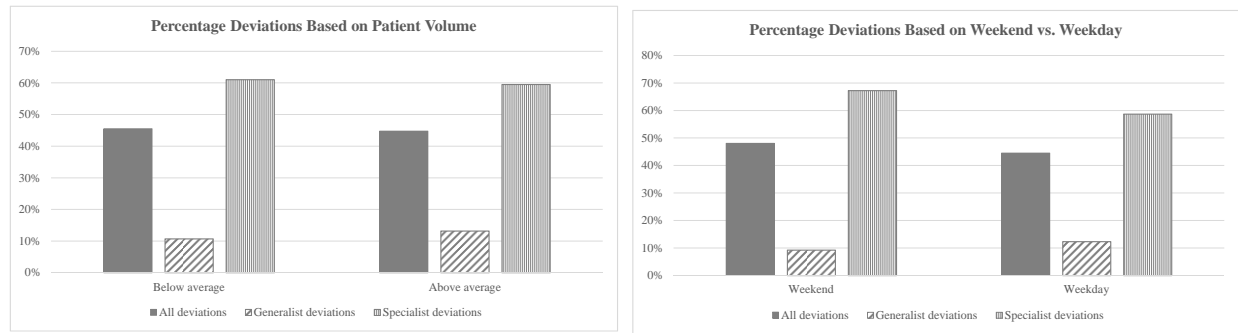
Notes: CCI = Chronic condition indicator. $n = 4,729$. * $p < 0.05$.

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 routine assignments occur more during (1) high volume periods (versus low volume periods), (2) weekdays (versus weekends), and (3) earlier shifts (versus later shifts)? The results are presented in Figure 1. Part (a) of this figure indicates that percentage deviations do not differ much 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 fairly 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.

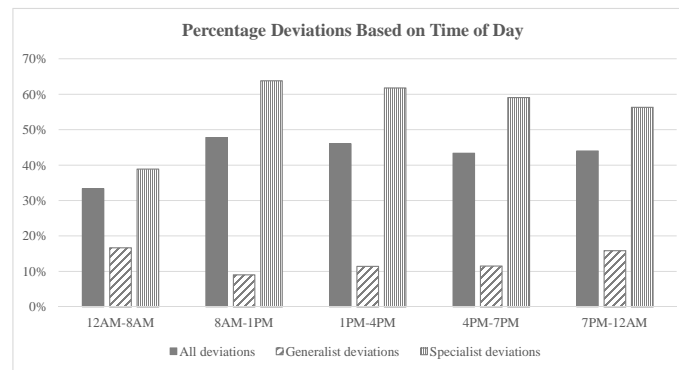
Next, we examined the effect of task type on routine assignment, as shown in Table 2. The analysis also included professional type (generalist or specialist) to further examine patterns in routine assignment that relate specifically to professional role that may be due to knowledge differences between the two, 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.

From Table 2, we observe the following. First, a generalist assignment had a positive and statistically significant estimated coefficient. The AME indicates that each additional hospitalization



(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

assigned to a generalist increases the likelihood of routine assignment by 48.2%; when compared to the sample average for routine assignment of 54.8%, this increase is actually 87.8%. By looking at task categories, we also observe from Table 2 that well-defined tasks had a statistically insignificant coefficient. However, tasks that were high in complexity, contact, and resource intensity each had statistically significant results. High complexity tasks had a positive coefficient ($p < 0.001$), and AME results indicating a 16.0% increase in the likelihood of routine assignment, or 29.2% 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 decrease the probability of routine assignment by 6.2%, or 11.3% when compared to the sample average of 54.8%. Resource-intensive tasks had a positive coefficient ($p < 0.001$), that represented an increase in the probability of routine assignment by 8.8%, or 16% against the sample average of 54.8%.

In Table 3, we report the performance implications of routine 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

Table 2 Task-Related Factors Associated with Routine Assignment

Dependent Variable: Routine Assignment	Coefficients	AME
Generalist (vs. Specialist) assigned	3.530*** (0.269)	0.4820
Well-defined task	-0.046 (0.101)	-0.0061
High complex task	1.237*** (0.117)	0.1602
High contact task	-0.469*** (0.136)	-0.0621
Resource-intensive task	0.387** (0.153)	0.0.0881

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. ** $p < 0.01$, *** $p < 0.001$. AME = Average marginal effect.

characteristics and year fixed effects, with standard errors clustered by physician. As M1 and M2 show, routine 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 routine 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 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 cost outcomes. However, M3 and M4 show that high complexity tasks, compared to those with low complexity, have readmission occurrences that were significantly lower ($p < 0.01$), 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 readmission and adverse event rates for resource intensive tasks ($p < 0.05$). Finally, Table 4 provides a summary of the results provided 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 routine 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 routine assignment will yield higher performance for tasks that are well-defined. The results in Table EC.1 (see M5 and M10 there) show that *well-defined* \times *routine assignment* interaction has a negative coefficient for both operational efficiency and cost outcomes ($p < 0.001$ for length

Table 3 Performance Implications of Routine Assignment and Task Categories

Dependent Variable: Variables	Operational Efficiency	Cost	Quality	
	Length of stay (days)	Total charges (\$)	Readmission, 30-day	Adverse event
	M1	M2	M3	M4
Routine assignment	0.173*** (0.035)	0.218*** (0.036)	0.259 (0.183)	0.200 (0.202)
Generalist (vs. Specialist assigned)	-0.280*** (0.085)	-0.565*** (0.097)	-0.183 (0.214)	-1.173*** (0.284)
Well-defined task	-0.003 (0.031)	-0.011 (0.032)	-0.758*** (0.136)	0.573*** (0.157)
High complexity task	-0.007 (0.034)	0.016 (0.035)	-0.451** (0.172)	-1.427*** (0.189)
High contact task	0.125** (0.042)	0.238*** (0.043)	-0.356 (0.190)	0.023 (0.191)
Resource-intensive task	0.102* (0.048)	0.135** (0.049)	-0.551* (0.236)	-0.697* (0.274)
Constant	1.448*** (0.082)	10.228*** (0.088)	-2.741*** (0.348)	-2.388*** (0.336)
No. of Observations	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	+	+	-	-

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

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 routine 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 routine and deviating assignments, we observe that routine assignments are associated with an increase of 0.466 days and \$8,474 in length of stay and total charges, respectively. 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 routine; and (2) although routine 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 indicate that routine assignment is more detrimental for poorly defined tasks (increase in length of stay by 2.18 days and in total charges by

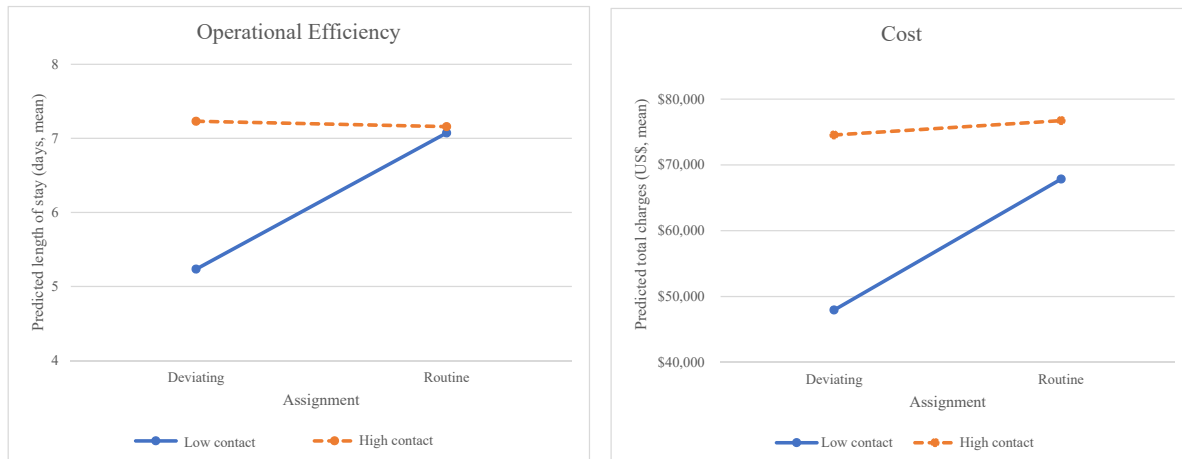


Figure 2 Predicted Performance Outcomes Resulting from Interaction of Routine Assignment and Well-Defined Tasks

\$22,700), 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 routine 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* \times *routine 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 routine assignment.

Hypothesis 3 states that routine 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 total charges have a statistically significant negative coefficient ($p < 0.001$) for the *high contact* \times *routine assignment* interaction. However, in terms of quality outcomes, M22 shows a statistically significant positive effect ($p < 0.05$) on adverse events pertaining this interaction, indicating that routinely assigned high contact tasks have an increased incidence of adverse events. However, our results show statistically insignificant effects in the other quality outcome, readmission rate, as shown in M17. To further evaluate these effects, Figures 3(a) and 3(b) demonstrate the predicted cost outcomes associated with *high contact* \times *routine assignment*. These graphs show partial support for our hypothesis, in that compared to deviations in assignment, routine assignment reduces length of stay by 0.072 days but increases total charges by \$2,178. 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 routine assignment for high contact tasks increases the likelihood of adverse events by 4.14%, compared to situations when assignment deviates. Thus, our findings are somewhat mixed for outcomes related to high contact tasks with routine assignment: while operational efficiency improves, cost and quality (in terms of adverse events) degrade.



(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 Routine Assignment and High Contact Tasks

Finally, Hypothesis 4 suggests that following the routine 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* \times *routine assignment*), Figures 4(a) and 4(b) provide graphs illustrating the moderating effect, which show that in comparison to deviations in assignment, routine assignment involves lower length of stay by approximately 2.25 days and lower total charges by \$23,566 when tasks have high resource intensity.

Summary of Hypothesis Tests. The summary of our results regarding our hypothesis tests is provided in Table 5. A quick look at this table shows that we can reject our first hypothesis, Hypothesis 1, with respect to operational efficiency and total cost outcomes. We can also reject



(a) Operations Outcome (Length of Stay)

(b) Cost Outcome (Total Charges)

Figure 4 Predicted Performance Outcomes Resulting from Interaction of Routine 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.

Hypothesis 3 with respect to operational efficiency. On the other hand, our results indicate that Hypothesis 3 is true with respect to total cost and adverse event outcomes, and Hypothesis 4 is true with respect to operational efficiency and total cost outcomes. Taken together, these results suggest that providing physicians with the flexibility to deviate from routine 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, routines should be enforced, as deviations either negatively impact performance or yield no tangible benefits.

4.1. Heterogeneity: Routine Type

In the previous section, we examined the implications of following routine assignments. Our results (see, e.g., Table 5) provide insights by considering the fact that the implications of following routine assignments might depend on (a) task type, and (b) the performance metric. To gain further insights into the implications of following routine assignments, we now also consider the fact that such implications might depend on the type of the provider that is assigned. Specifically, we note that the impact of following routine assignment might be heterogenous among the two types of

Table 6 Routine Type Heterogeneity (Summary Results)

Hypothesis #	Operational Efficiency		Cost		Quality			
	Length of stay (days)	Length of stay (days)	Total charges (\$)	Total charges (\$)	Readm. 30-day	Readm. 30-day	Adverse event	Adverse event
	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	T	F	T	F	-	-	F	T
2	-	-	-	-	-	-	-	T
3	F	T	F	T	-	-	-	T
4	T	T	-	T	-	-	-	T

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

routine: following routine when a generalist is assigned (labeled as Type I routine) and following routine when a specialist is assigned (labeled as Type II routine).

In Table 6, we provide the summary of our hypothesis tests by making use of hospitalizations with Type I and Type II routines separately. Our results show that, as is expected, there is some amount of heterogeneity in our results. Specifically, focusing on hypotheses and performance metrics for which considering Type I and Type II routine assignments separately would alter the results of our hypothesis tests (i.e., from true to false or from false to true), we observe two notable changes in Table 6 compared to Table 5: (1) Among Type I routines, Hypothesis 1 becomes true with respect to length of stay and total charges metrics, and Hypothesis 3 becomes false with respect to total charges. (2) Among Type II routines, Hypothesis 3 becomes true with respect to the length of stay measure. These suggest that (a) length of stay and total charges metrics are sensitive to the type of the routine, and (b) this sensitivity is especially higher for well-defined tasks (Hypothesis 1) and high contact tasks (Hypothesis 3). Overall, we find that the implications of our results are not completely homogenous and depend on the provider type that is assigned in practice. This is, however, to some extent expected; intuitively, assigning a generalist when a specialist should have been assigned has different consequences than assigning a specialist when a generalist should have been in charge.

4.2. Mechanism: Mediation Analysis

In order 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 routine 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 routine 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.

Results 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 high contact tasks, we find that advanced imaging has a partial mediation effect for length of stay, and total charges. Employing bootstrapping analysis with 1,000 iterations, routine assignment for high contact tasks shows a significant effect on advanced imaging ($B = -0.454$, $se = 0.126$, $p < 0.001$), which in turn has a significant effect on length of stay ($B = 1.105$, $se = 0.020$, $p < 0.001$) and total charges ($B = 101,134$, $se = 16,970$, $p < 0.001$). For these outcomes, the natural direct effect (NDE) and natural indirect effect (NIE) are significant, indicating partial mediation. In terms of resource-intensive tasks, routine assignment demonstrates a significant effect on advanced imaging ($B = -0.274$, $se = 0.120$, $p = 0.022$) which subsequently has a significant effect on total charges ($B = 90,062$, $se = 15,636$, $p < 0.001$). In the case of resource intensive tasks, the NDE is not significant while the NIE is significant, suggesting that advanced imaging fully mediates the relationship between routine assignment and total charges. We do not observe mediation effects for well-defined and 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 routine 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 ($B = 53,474$, $se = 4,663$, $p < 0.001$). For these outcomes, the NDE and NIE, and total effect are all significant, indicating partial mediation. Further, the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero for advanced imaging, suggesting significant mediation. 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 ($B = 81,019$, $se = 9,453$, $p < 0.001$). For these

Table 7 Predicted Impact of Different Counterfactual Scenarios (Compared to Current Practice) Due to Adhering to Routine 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.

outcomes, the NDE, NIE, and total effect are significant, indicating partial mediation, which is also supported by the NIE 95% bias-corrected confidence interval excluding zero. However, we again do not observe significant mediation 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 fully mediates the relationship between routine assignment and total charges for resource-intensive tasks. In addition, we find that advanced imaging partially mediates the relationship between following the routine assignment and outcome variables such as length of stay and total charges, especially for high contact tasks. The strength of the mediating impact of use of advanced imaging also depends on whether the routine 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.

4.3. Providing Recommendations Using Cost-Effectiveness Analyses

To provide clear recommendations for hospital administrators if they choose to develop more formalized guidelines around routine assignments, we now consider five counterfactual policies corresponding to imposing the guideline fully or partially. Specifically, in Table 7 we summarize the predicted impact of adhering to routine 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 the 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 4.3, 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

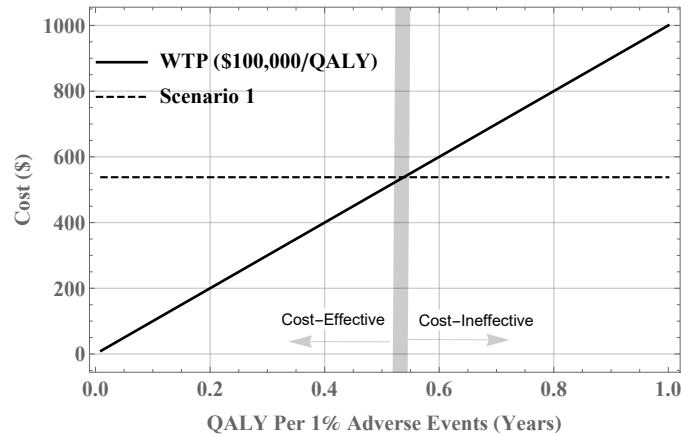


Figure 5 Cost-Effectiveness of Scenario 1 (Imposing the Guideline for All Task Types).

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).⁶ 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 scenario among the five counterfactual ones considered, as it yields tangible benefits without any significant impact on quality metrics.

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 approach with specifications comparable to our original GLM models. Specifically, we fit GLM models using 2-stage nonlinear least squares. Our selected instrument is whether a generalist versus specialist was assigned as we observe that this variable is (a) highly correlated with assignment matching guideline ($corr = 0.44$, $p < 0.001$), and (b) is not correlated with the four outcomes of interest. Further, we find an F-statistic of 435 ($p < 0.001$) for the first-stage estimator, which surpasses the minimum threshold F-statistic of 10

⁶ 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.

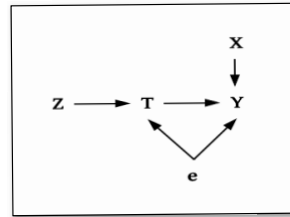


Figure 6 The Directed Acyclic Graph (DAG) Behind Our Instrumental Variable Approach [Z : generalist is assigned; T : assignment matches the guideline; Y : one of the four outcome variables; X : vector of control variables].

typically required for identifying a weak instrument (see, e.g., Stock et al. 2002). Both of these tests provide support that our instrument is not weak.

We conceptualize our approach using the Directed Acyclic Graph (DAG) depicted in Figure 6. In this figure, the treatment variable T represents when assignment matched the guideline (routine 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 any of our outcome variables (length of stay, total charges, 30-day readmission, and adverse events), 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). Similar to our prior GLM specifications, we use a gamma family with log link for the length of stay and total charges outcomes, and binomial family with logit link on 30-day readmission and adverse events outcomes. We also cluster each model by physician, as we did previously.

Our results are presented in Table 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 routine assignment on performance outcomes is consistent with our earlier results across all four task categories and four outcome variables. 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, gives us confidence that our results 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 routine/guideline) and control (i.e., assignment deviating from routine/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

Table 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 9 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.

Table 10 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

robustness). The results are provided in Tables 9 and 10. As can be seen from Table 9, the results 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 10 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 routine 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. 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 percent agreement to 45% and 55% instead of 50%. Tables EC.4 and EC.5 (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 to define consensus opinion for each medical diagnosis in the survey. Table EC.6 (see the Appendix) displays similar results to our earlier reported findings, supporting the robustness of our results to our original measure for consensus opinion.

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

Finally, we assessed the robustness of our results to the assumption that our outcome variables are independent. 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, of course, 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 the 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 routine assignment guidelines did not formally occur in our analysis. Specifically, we built our understanding of routine assignment on prior work explaining how routines 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 routine practice. In the ideal study, which future research in this area could attempt, routinized practices would be formally made known as guidelines to professionals, and analysis of performance could occur using prospective data collected from the implementation of routines onward. 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 would have involved a scheduled hospitalization in the future 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 endogeneity issues, 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 randomized control trial (RCT), which can further test the validity of the evidence we establish.

7. Conclusion

We study the impact of professionals using routine versus deviating task assignments on various performance outcomes, and shed light on whether enforcing formalized guidelines reflecting these routine assignments can boost performance. Our work is motivated by the contrast between (a) operations management and organizational theories that promote standardization and reduction of use of discretionary judgements, and (b) decision theories and the conventional wisdom that suggest professionals should be given the opportunity to use their discretionary judgement and deviate from routine assignments when needed. Our results allow a *task-type* view of this contrast by generating insights into specific task types for which permitting deviations can improve performance.

To perform our analysis, we use data on the assignments of generalist and specialists to patients in a children’s hospital, and consider three sets of metrics: operational efficiency (measured by length of stay), quality of care (measured by occurrence of 30-day readmission and adverse events), and cost (measured by total charges). Specifically, we make use of detailed patient-level EMR data of more than 4,700 hospitalized patients who had one of the 73 top medical diagnoses. Our EMR data includes various patient demographic information, the nature of the diagnoses, patient outcome measures, and the physician ultimately taking responsibility for the patient’s care. In parallel, we conduct detailed physician surveys and utilize it to identify the type of provider that should have been in charge of each patient based on the medical diagnosis of the patient.

Comparing these two sources of data, we find that providing physicians with the flexibility to deviate from routine 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 routine assignments. These insights provide a better understanding of the underlying tradeoffs in standardizing assignment decisions through routines, and can help hospital administrators balance permitting and not permitting use of discretionary judgements in their practice. 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 are beneficial from some aspects, they are simultaneously detrimental from other aspects. Hence, in contrast with the finding that enforcing routine assignments is the dominant strategy for some tasks types (e.g., highly complex or resource-intensive tasks), permitting deviations is typically not dominantly the better option (regardless of the task type). In addition, our results indicate that there may be an intriguing interplay between professional status and discretionary deviations from routine 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. Of note, these evidence on no free lunch and assertion of power might be specific to our setting (similar to some of our other results), and generalizations to other hospitals or care provided in non-pediatric environments without further tests should be avoided.

Finally, our results show that there might be environmental conditions under which deviations are invoked more frequently. In particular, we find that such deviations are invoked more during weekends than weekdays, and during morning shifts (8am-1pm) than other shifts. However, we observe that deviations are invoked 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 routine assignments. Future research can also go beyond the healthcare sector, and investigate other professional settings in which a task-type view can inform the advantages and disadvantages of disallowing use of discretionary judgements in deviating from routine 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.
- 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.
- Becker, Markus C. 2004. Organizational routines: A review of the literature. *Industrial and Corporate Change* **13**(4) 643–678.
- 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.
- Black, N., M. Murphy, D. Lamping, M. McKee, C. Sanderson, J. Askham, T. Marteau. 1999. Consensus development methods: A review of best practice in creating clinical guidelines. *Journal of Health Services Research & Policy* **4**(4) 236–248.
- Bodenreider, Olivier, Barry Smith, Anita Burgun. 2004. *The ontology-epistemology divide: A case study in medical terminology*. IOS Press, 185.
- 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.
- Boone, Christophe, Vera Brdcheler, Glenn R. Carrol. 2000. Custom service: Application and tests of resource-partitioning theory among dutch auditing firms from 1896 to 1992. *Organization Studies* **21**(2) 355–381.
- Borgatta, Edgar F., Robert F. Bales. 1953. Task and accumulation of experience as factors in the interaction of small groups. *Sociometry* **16**(3) 239–252.
- 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.
- 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.
- Entin, Elliot E., Daniel Serfaty. 1989. Adaptive team coordination. *Human Factors* **41**(2) 312–325.
- Feldman, Martha S., Brian T. Pentland. 2003. Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly* **48**(1) 94–118.
- Feldman, Martha S., Anat Rafaeli. 2002. Organizational routines as sources of connections and understandings. *Journal of Management Studies* **39**(3) 309–331.
- 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.
- Fong Boh, Wai, Sandra A. Slaughter, Alberto J. Espinosa. 2007. Learning from experience in software development: A multilevel analysis. *Management Science* **53**(8) 1315–1331.
- Freidson, Eliot. 1970. *Dominant Professions, Bureaucracy, and Client Services*. Merrill Publishing Co., 428–448.
- 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.
- Greenwood, B. N., R. Agarwal, R. Agarwal, A. Gopal. 2019. The role of individual and organizational expertise in the adoption of new practices. *Organization Science* **30**(1) 191–213.
- 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.
- Hong, Bryan, Lorenz Kueng, Mu-Jeung Yang. 2019. Complementarity of performance pay and task allocation. *Management Science (forthcoming)*.
- Hopp, W.J., S.M.R. Irvani, 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.
- Kahn, Robert L., Donald M. Wolfe, Robert P. Quinn, J. Diedrick Snook, Robert A. Rosenthal. 1964. *Organizational Stress: Studies in Role Conflict and Ambiguity*. John Wiley & Sons.
- 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.

- Kuntz, Ludwig, Stefan Scholtes, Sandra Sülz. 2016. Separate & concentrate: Accounting for process uncertainty in the design of regional hospital systems. *Working Paper, Columbia University*.
- 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.
- Murphy, M. K., N. A. Black, D. L. Lamping, C. M. McKee, C. F. Sanderson, J. Askham, T. Marteau. 1998. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment (Winchester, England)* **2**(3) i-88.
- Nelson, C.R., Sidney Winter. 1982. *Organizational Capabilities and Behavior: An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press.
- Orasanu, J.M. 1993. *Cockpit Resource Management*. Academic Press, 137-172.
- Orasanu, J.M. 2001. *New Trends in Cooperative Activities: Understanding System Dynamics in Complex Environments*. Human Factors and Ergonomics Society, 242-258.
- 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.
- 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 Chilingirian. 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.
- 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 production 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.
- Spee, Paul, Paula Jarzabkowski, Michael Smets. 2016. The influence of routine interdependence and skillful accomplishment on the coordination of standardizing and customizing. *Organization Science* **27**(3) 759-781.
- 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.
- Stinchcombe, Arthur L. 2001. *When Formality Works: Authority and Abstraction in Law and Organizations*. University of Chicago Press.
- 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.
- Timmermans, Stefan, Alison Angell. 2001. Evidence-based medicine, clinical uncertainty, and learning to doctor. *Journal of Health and Social Behavior* **42**(4) 342-359.
- Traub, S.J., C. Stewart, R. Didehban, A. Bartley, S. Saghafian. 2016. Emergency department rotational patient assignment. *Annals of Emergency Medicine* **67**(2) 206-215.
- Treem, Jeffrey W. 2012. Communicating expertise: Knowledge performances in professional-service firms. *Communication Monographs* **79**(1) 23-47.
- 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 Ven, Andrew H., Andre L. Delbecq, Richard Koenig Jr. 1976. Determinants of coordination modes within organizations. *American Sociological Review* **41**(2) 322-338.
- 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.
- von Hippel, Eric. 1990. Task partitioning: An innovation process variable. *Research Policy* **19**(5) 407-418.
- Waring, Justin, Graeme Currie. 2009. Managing expert knowledge: Organizational challenges and managerial futures for the uk medical profession. *Organization Studies* **30**(7) 755-778.
- Woolf, S. H. 1992. Practice guidelines, a new reality in medicine: Ii. methods of developing guidelines. *Archives of Internal Medicine* **152**(5) 946-952.