Public Relative Performance Feedback in Complex Service Systems: Improving Productivity through the Adoption of Best Practices

Hummy Song, Harvard University

Anita Tucker, Economics Department, Brandeis University

Karen L. Murrell & David R. Vinson, Kaiser Permanente

Working Paper Series

2015|96

Public Relative Performance Feedback in Complex Service Systems: Improving Productivity through the Adoption of Best Practices

Hummy Song* — Harvard University
Anita L. Tucker — Brandeis University
Karen L. Murrell and David R. Vinson — Kaiser Permanente

October 8, 2015

Abstract

Managers of service organizations seek to improve productivity without eroding service quality. We explore whether privately versus publicly disclosing relative performance feedback (RPF) about individual workers' processing times can help achieve this goal. Using three years of patient encounter data from two emergency departments, one of which changed from privately to publicly disclosing RPF to physicians, we find an 8.6% improvement in productivity and no significant reduction in quality associated with implementing public RPF. This benefit is greater when workers are carrying out unstandardized, rather than standardized, tasks. We conduct further analyses that suggest the benefit of public RPF may primarily stem from the identification and diffusion of best practices around workflow, rather than from the motivation to be top-ranked or the shame of being bottom-ranked. Thus, our results suggest public RPF can foster the sharing and adoption of strategies for improving the management of workflow.

Key words: productivity, workflow, relative performance feedback, best practice transfer, empirical operations.

*Corresponding author. Email: hsong@hbs.edu. Address: Wyss House, Soldiers Field Road, Boston, MA 02163.

1 Introduction

In service organizations, a key objective for managers is to create conditions that help workers improve their productivity without eroding service quality. One commonly employed approach for improving quality and productivity is standardized work (Spear and Bowen, 1999). This approach has been shown to positively affect quality and productivity in a variety of industries, including health care (Andritsos and Tang, 2014; Chandrasekaran et al., 2012; Shah et al., 2008), automotives (Ohno, 1988), retail (DeHoratius and Raman, 2008), banking (Frei et al., 1999; Staats and Gino, 2012), and software (Narayanan et al., 2009; Staats et al., 2011). In each of these cases, standardized work leads to operational improvements by reducing variability in how tasks are executed. For example, checklists ensure that workers do not forget key steps or deviate from evidence-based treatments (Pronovost et al., 2006).

Though the benefits of standardized work have been well documented, many processes in complex service organizations remain unstandardized, due to the challenges presented by inherent variability in customer needs and differences in worker behaviors that arise from worker discretion and imperfect monitoring (Klassen and Menor, 2007; Tucker et al., 2007). In such settings, it may not be feasible to develop and implement a standardized protocol for each specific work task. Instead, managers may need to consider alternate ways to motivate and enable workers to better manage their *workflow* across all tasks. We draw on Georgakopoulos et al. (1995) to define workflow as a worker's selection and sequencing of a set of tasks necessary to accomplish the objectives for his or her customers. To illustrate, standardized work would specify that emergency department (ED) physicians order a beta blocker within five minutes of diagnosing a patient with a heart attack. In contrast, standardized workflow would specify that physicians order at the beginning of a patient's visit all pertinent tests and medications needed to treat the most probable diagnosis, rather than ordering them in a serial fashion.

One way to standardize workflow may be by publicly disclosing—among workers within an organization—relative performance feedback (RPF) on productivity metrics. Note, for brevity, we do not always specify that we consider cases in which RPF is provided on productivity metrics (as opposed to other performance metrics, such as quality), but this should be assumed. RPF compares an individual worker's performance to his or her peers within the organization (Blanes i Vidal and Nossol, 2011). Nearly one-third of U.S. corporations provide RPF to their workers (McGregor, 2006; Nordstrom et al., 1991). We distinguish RPF, which is feedback provided specifically to the members within an organization, from public reporting, which is performance data made available to those external to an organization. For a review of public reporting in health care settings, see Fung et al. (2008).

When RPF on a productivity metric is disclosed *publicly* among workers within an organization (i.e., with each worker's name displayed alongside his or her productivity metric), workers may be able to determine how to improve their productivity by identifying high-productivity coworkers who have developed best practices around workflow. In addition, workers may also be more motivated to improve their productivity because public RPF makes performance differences across workers more salient. This is in contrast to when RPF is disclosed *privately*, in which the same information about relative performance is provided but the identities of coworkers are kept anonymous. The potential benefits of public RPF on improving worker productivity may be especially relevant for complex service systems in which standardized work is not in place, due to the resulting variation in how workers approach their work. Despite the frequent use of public RPF, scant research has investigated its impact on worker productivity. Therefore, we ask the following research questions: *What is the effect of public RPF on worker productivity? Does this effect vary by whether the work is standardized, and what are the implications for service quality? What are the mechanisms through which public RPF affects worker productivity?*

In this paper, we explore these questions in the context of a hospital ED, which is a complex service organization with significant variability in patients' needs and inherent variability in processing times across physicians due to their ability to exercise discretion in carrying out their work (McCarthy et al., 2012). Here, certain work tasks are standardized, but many are not. For example, for patients presenting with symptoms indicative of a stroke, there are clear protocols for diagnosis and treatment in the ED. However, for patients presenting with abdominal pain, which could be indicative of a high severity condition, standardized protocols are typically not in place. Improvements in the management of workflow and associated reductions in variability are expected to have a significant positive effect on operational performance in this setting (Soremekun et al., 2011). Such improvements are particularly needed in EDs given the high demand for ED care (Pitts et al., 2010) that coexists with declining reimbursement rates (Morganti et al., 2013) and high rates of uncompensated care (Centers for Medicare & Medicaid Services, 2002).

For our identification strategy, we leverage an exogenous change to the way RPF is provided to physicians at one of two EDs within the same health care system. During the initial time period of our study, physicians at both EDs were provided private RPF in the form of a ranked histogram about the median ED length of stay (LOS) of their patients—an important productivity metric in an ED setting. With this private RPF, each physician's identity remained anonymous and was represented with a code number that only he or she knew. Beginning in August 2010, physicians at one ED (but not the other) received the same RPF information in a public manner where each physician's name was listed next to his or her median LOS on the ranked histogram. Thus, with public RPF, physicians were able to identify high-productivity peers who could share best practices regarding workflow. Using a difference-in-differences approach, we first estimate the impact on worker productivity of changing from private to public RPF. We then assess how this effect varies by whether the patient's condition has a standardized work protocol. In addition, we examine the effects of this change on service quality. Finally, assuming that the hypothesized positive impact of public RPF on productivity exists, we consider three mechanisms that may explain this effect: motivation to be at the top of the performance distribution, motivation to avoid the shame of being at the bottom of the performance distribution, and the identification and diffusion of best practices around workflow.

Our results show that, on average, public RPF is associated with an 8.6% increase in physician productivity. We find that the increase in productivity is not statistically significant for patient conditions with standardized processes but *is* statistically significant for patient conditions without standardized processes. In addition, we find a slight decrease in the amount of care provided to patients and no significant reduction in clinical quality or patient satisfaction. Based on our analyses of heterogeneous treatment effects and changes in the variation in processing times, the explanation that is most consistently supported for why public RPF leads to improved productivity is that public RPF enables the identification and diffusion of best practices around workflow. Data from physician interviews and observations reinforce this conclusion.

This paper makes several contributions to both theory and practice. First, we contribute to the operations management literature on feedback (e.g., Bendoly, 2013; Schultz et al., 1999) by examining the effects of public RPF in a field setting. To our knowledge, there have been few studies of public RPF, and most have been conducted in laboratory settings. This line of research is important both theoretically and practically because feedback provides an additional lever for influencing worker behavior that compliments operations management's typical focus on standardized work. Second, we illustrate that focusing on improving the management of *workflow*, as opposed to the standardization of *work tasks*, may be particularly useful when workers have discretion in how to carry out their work (Hopp et al., 2009). In doing so, we build on the operations literature that examines workflow management in health care settings (Dobson et al., 2013; KC, 2014). Third, we illuminate conditions under which public RPF is most likely to lead to improved productivity. We identify the importance of providing such feedback on metrics that can be improved by the spread of best practices rather than simply reflective of differences in individual ability. We also find that providing public RPF on productivity metrics is more likely to result in productivity gains when workers are carrying out unstandardized work tasks as opposed to ones that are already standardized. Finally, we provide insight into the mechanisms through which public RPF has a positive impact on productivity.

2 Related Literature and Hypotheses

2.1 Productivity Improvement in Service Organizations

Operations management scholars have examined various approaches to improve productivity at the level of the individual worker. One commonly employed approach is standardized work (Spear and Bowen, 1999), which has been discussed in the operations management literature since Taylor's (1911) seminal research in this area. Standardized work reduces process variability, which in turn helps improve productivity by reducing customer waiting times (Hopp and Spearman, 2000; Kingman, 1961) and facilitating learning through process control (Bohn, 1995; Jaikumar, 2005). The positive effects of standardized work on quality and productivity have been documented in studies spanning several industries, including health care (Andritsos and Tang, 2014; Chandrasekaran et al., 2012; Shah et al., 2008), automotive (Ohno, 1988), retail (DeHoratius and Raman, 2008), banking (Frei et al., 1999; Staats and Gino, 2012), and software (Narayanan et al., 2009; Staats et al., 2011).

Of course, standardized work only helps improve productivity and quality if employees are compliant with the specified procedure. Research suggests that compliance with established procedures is often low, regardless of industry (Anand et al., 2012; Staats et al., 2015). For example, in pharmaceutical manufacturing plants, noncompliance with operational routines is identified in more than half of all inspections conducted by the Food and Drug Administration (Anand et al., 2012). Similarly, despite it being widely established that handwashing leads to improved patient outcomes, average compliance rates with handwashing guidelines are documented to be consistently below 50% (Centers for Disease Control and Prevention, 2002). Electronic monitoring has been examined as a potential way to ensure greater compliance (Pierce et al., 2015; Staats et al., 2015), but this is not without challenges. In a paper examining the effectiveness of electronic monitoring to encourage greater hand hygiene compliance, Staats et al. (2015) finds that compliance rates fall *below* pre-monitoring levels after electronic monitoring is discontinued. This suggests that monitoring efforts should be implemented with care and sustained managerial commitment.

The many benefits of standardized work suggest that managers should strive to develop, implement, and ensure compliance with standardized work. However, in complex service systems with significant heterogeneity in customer needs, standardization may not always be feasible to develop and implement at the level of each specific customer type. Furthermore, service workers often have to prioritize among different customers who need service simultaneously, making it difficult to develop absolute rules about which tasks to perform and in which order (KC, 2014; Tucker and Spear, 2006; Wang et al., 2015). In such settings, it may be helpful to consider instead how workers can best manage their *workflow* across all their tasks. One approach may be to select the appropriate level of multitasking that allows for faster service rates without sacrificing quality (KC, 2014). Prior research on service workers has found that workers who perform a variety of tasks enjoy productivity benefits over employees with repetitive tasks (Narayanan et al., 2009; Staats and Gino, 2012). Another approach may be to manage interruptions by adhering to optimal prioritization policies between new customers and those already in the system (Dobson et al., 2013).

Given the context-dependent nature of these decisions around the management of workflow, an effective approach to identifying best practices in a given setting may be to have the frontline workers define these approaches themselves. Yet, with each worker employing a different set of practices, which are not always fully transparent to others, these practices may be difficult to identify and isolate. Furthermore, employees might be inclined to adhere to their own approaches, even if their methods are non-optimal (Berwick, 2003). How might managers help motivate and enable workers to better manage their workflow across all tasks?

2.2 Relative Performance Feedback in Service Organizations

One way to help identify best practices around workflow—especially those that are developed by the frontline workers themselves—may begin with identifying the most productive workers in an organization. Managers can enable workers to identify their most productive peers by publicly disclosing individual-level RPF on productivity metrics among workers within an organization. By identifying these workers and their approaches for managing workflow, other workers within the organization may be able to identify a set of best practices to adopt and be motivated to change their practices around workflow.

Even when privately disclosed, RPF leverages the effects of social comparison by providing workers with visibility into the full distribution of all workers' performance. Social comparison has been shown to have powerful effects on worker behaviors. Some workers find social comparisons to be motivating, especially when monetary incentives are linked to relative performance. When provided social comparisons, workers tend to increase their productivity, as has been shown by prior studies in both laboratory (Charness et al., 2014; Kuhnen and Tymula, 2012) and field settings (Blanes i Vidal and Nossol, 2011; Cowgill, 2015). However, social comparisons can lead to negative effects as well, in which workers exert less effort as opposed to more (Ashraf et al., 2014; Barankay, 2012). For example, in a field study of furniture salespeople, Barankay (2012) finds that low-ranked workers exert less effort when presented with feedback on rank due to the effect of a low ranking on an individual's self-image. In another field study with individuals training to become Zambian health workers, Ashraf et al. (2014) finds that social comparisons negatively affect the performance of low-performing trainees, as measured by their test scores on a training course exam. Perhaps even worse, some research finds that workers receiving unfavorable social comparisons are likely to engage in deceptive behaviors to artificially inflate their reported performance (Edelman and Larkin, 2015; Moran and Schweitzer, 2008). Although this may improve a worker's reported relative performance measure, it has negative implications for the organization's performance and trust among coworkers (Dunn et al., 2012).

Because of these varying potential implications, care is needed when deciding how to implement RPF. First, RPF can be disclosed either privately or publicly amongst the workers. With private RPF, a worker cannot identify which individual coworker exhibited a specific performance level, and the worker knows that likewise her coworkers cannot link her own performance result to her. In contrast, with public RPF, a worker can link each performance level to an individual coworker and knows that her coworkers can also see how she herself has performed. Second, managers can link monetary incentives and compensation to relative performance. Perhaps unsurprisingly, many studies find that RPF motivates workers to work harder when monetary incentives are linked to relative performance (Tafkov, 2013). Some studies also find that, even without monetary incentives, RPF can lead to increases in productivity by stoking intrinsic competition (Blanes i Vidal and Nossol, 2011; Roels and Su, 2014). In this paper, we focus on privately versus publicly disclosing RPF without any monetary incentives for productivity. If this approach is successful at increasing productivity, it may be easier and more cost-effective to implement than having to financially incentivize workers for their productivity.

2.3 Private Versus Public Relative Performance Feedback

Much of the prior research on RPF has focused on examining the effects of *private* RPF, either in the presence or absence of monetary incentives. To our knowledge, Tafkov (2013) and Hannan et al. (2013) are the only studies that examine the relationship between public RPF and worker performance. In a laboratory setting with participants solving multiplication problems for pay, Tafkov (2013) finds that participants solve more problems correctly when they are provided public as opposed to private RPF. Although this study finds improvements in performance with public RPF, it is unclear whether such effects would hold in a field setting with complex tasks, various approaches for carrying out these tasks, and workers who know one another. Hannan et al. (2013) builds on these results in a multi-task environment and shows that when workers can allocate their time across two different types of tasks, RPF leads workers to distort their time allocation away from the optimal allocation (50/50 in the experiment) to favor the task on which they performed better in an earlier round. The authors find that the effort distortion effect is stronger with public RPF than private RPF, and persists even if it decreases their remuneration for the experiment. These studies highlight that public RPF provides strong motivation for workers to change their behaviors.

In this paper, we examine how public RPF affects workers' productivity, on average, compared to private RPF when workers are operating in a complex service organization. From the perspective of the worker, there are two key differences between public versus private RPF: (a) the knowledge that coworkers can see his or her own level of performance, and (b) the knowledge of each coworker's level of performance. The first of these two elements may motivate workers to increase their productivity more than when provided private RPF, due to a heightened level of social pressure (Mas and Moretti, 2009; Schultz et al., 1999). Receiving public RPF is likely to have a greater effect on productivity than private RPF due to the increased salience of being monitored (Staats et al., 2015). This effect manifests not only when the worker's productivity level is visually observable to others (Schultz et al., 1999) or to one's high-performing peers in particular (Mas and Moretti, 2009), but at all times. The second element may equip workers with the necessary knowledge to identify top performers and their best practices around workflow in a credible way, which may further improve workers' productivity on average. This leads us to our first hypothesis:

Hypothesis 1 (H1): Public RPF leads to a decrease in processing time, on average.

Assuming that public RPF does lead to an improvement in worker productivity, we are interested in assessing whether this productivity gain is particularly salient when standardized work is not in place and there is variation in how workers approach their work. This helps to determine whether providing public RPF may be a useful tool for improving productivity in complex service systems in which standardized work is difficult to implement—the type of setting in which we are most interested in this study.

Take for example a hospital ED. In this setting, there is no standardized work protocol for the diagnosis and treatment of patients presenting with symptoms of abdominal pain. Physicians develop their own workflow that fits their own needs and the needs of the other patients under their care. This results in significant variation in workflow across physicians, leading to lower average worker productivity than would occur if all workers employed the best practices of their high-productivity peers.

At the same time, some medical conditions have standardized work protocols. For example, there is a clear protocol for diagnosis and treatment in the ED for patients presenting with symptoms indicative of a heart attack or stroke. When processes for standardized work are already in place, there may not be much room for improvement from adopting the best practices of high-productivity peers. Thus, we expect public RPF to be particularly helpful in improving productivity when standardized work is not in place:

Hypothesis 2 (H2): The decrease in processing time resulting from a shift from private to public RPF is greater when standardized processes are not in place compared to when they are.

The hypothesized benefit of public RPF on worker productivity would be called into question if it were accompanied by a decrease in quality. After all, managers seek to improve productivity *without* eroding service quality. Consequently, we also examine, but do not hypothesize about, the impact of public RPF on service quality.

There are several potential mechanisms that may result in increased productivity when switching from private to public RPF. First, workers may be motivated to be at the top of the relative performance distribution. Past research suggests that individuals care about the prestige of being a top-ranked performer (Charness et al., 2014). Furthermore, this concern about high rank is innate (Zizzo, 2002). Given the greater utility experienced when attaining a higher rank on a relative performance distribution, workers who have access to public RPF may exhibit ahead-seeking behaviors, which is when they seek to outperform others by increasing their own performance (Roels and Su, 2014). Workers who are already toward the top of the distribution are most likely to exhibit such behaviors, since the additional effort required to remain in the top is modest compared to what a mid- or bottom-performing worker would need to exert (Boudreau et al., 2015). That said, it is possible that motivation to be at the top is no stronger when RPF is publicly as opposed to privately disclosed. This may be because workers are intrinsically motivated to be at the top even if their peers are ignorant of their high level of performance. Furthermore, workers may care about their ranking even if it remains private because their managers know how their performance compares with their peers' performance (Nagin et al., 2002).

A second potential mechanism is workers' motivation to avoid being at the bottom of the relative performance distribution. Bottom-ranked workers may exert greater effort and seek to improve their productivity in order to avoid the shame or embarrassment of being in last place. When RPF is disclosed publicly as opposed to privately, last-place aversion becomes more salient for these workers, as their bottom-ranked status becomes disclosed to others (Kuziemko et al., 2014). Nevertheless, it is also possible that these workers become discouraged by upward social comparisons, as has been documented in the context of retirement savings (Beshears and Gino, 2015), health worker trainings (Ashraf et al., 2014), fruit picking (Bandiera et al., 2013), and furniture sales (Barankay, 2012). Thus, it is unclear whether bottom-ranked workers will exert more or less effort to move up in the rankings, and whether the motivation will be stronger with public or private RPF.

A third potential mechanism is the identification and diffusion of best practices that is enabled by public RPF. Public RPF makes the top performers easily identifiable, which in turn enables workers to learn best practices either by observing how the top performers work or by directly asking them to share their productivity tips. Public RPF also helps overcome an important barrier to best practice diffusion, which is the perceived credibility of the information. When best practices are shared while RPF is privately disclosed, the "anonymous" information may lack the credibility necessary to change behaviors. Public RPF thus reduces barriers to both the identification and diffusion of best practices, enabling and motivating workers to seek out reliable ways to improve productivity (Szulanski, 1996). This identification of top performers and their best practices enables peer-based learning, which has been found to be more important for productivity improvement than learning by doing (Chan et al., 2014). In a study of department store salespeople, Chan et al. (2014) finds that workers learn from each other through direct observation and active teaching when they are working on the same shift alongside one another. Similarly, Kuziemko et al. (2014) finds that top-ranked workers are willing to engage in behaviors to help low-ranked people improve their performance.

These three potential mechanisms need not be mutually exclusive of one another in explaining how public RPF leads to improvements in productivity. In fact, it is likely that each of these mechanisms contributes to some extent to the overall effect of public RPF on worker productivity. While our primary objective in this study is to examine the effect of public RPF on worker productivity and the degree to which this effect varies by whether standardized work is already in place, we also explore which of these mechanisms may serve as the leading explanation for what may be driving the effects of public RPF on worker productivity.

3 Setting and Data

3.1 Research Setting

We examine our research questions in the context of a hospital ED. In this setting, there is significant variability in patient needs and inherent variability in processing times across physicians (McCarthy et al., 2012). In addition, while certain work tasks are standardized, many are not.

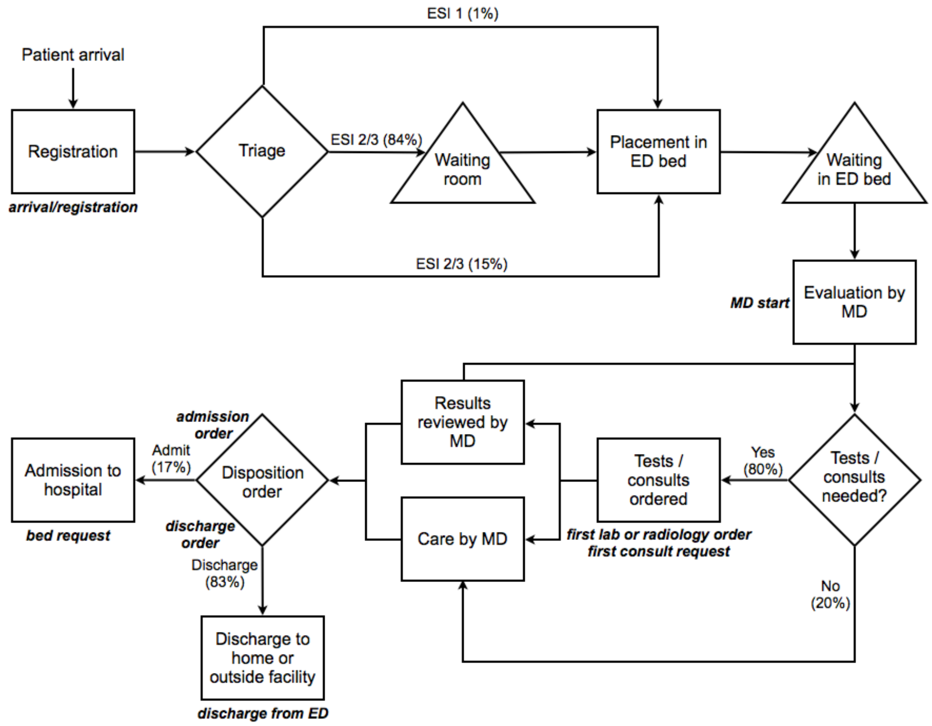
The ED is a setting in which a key objective is to attain shorter waiting times and LOS while simultaneously minimizing adverse outcomes (McClelland et al., 2011). When LOSs are longer, EDs become crowded, which prevents new arrivals from being seen in a timely manner (i.e., longer waiting times), and ultimately increases the risk of adverse patient outcomes (Guttmann et al., 2011; Schull et al., 2015). In addition, longer waiting times and LOS are associated with lower levels of patient satisfaction (Spaite et al., 2002).

Figure 1 is a schematic representation of a typical patient journey through the ED. A patient arrives either via ambulance or as a walk-in, at which point an ED clerk registers the patient. A triage nurse then obtains vital signs, collects the chief complaint, and assigns a 5-level Emergency Severity Index (ESI) triage category with 1 being the most urgent and 5 being the least urgent (Gilboy et al., 2011). Unless the patient is of ESI level 1 (in which case the patient is immediately taken to the resuscitation room) or there is an available ED bed, the patient returns to the waiting room until an ED bed becomes available. Once the patient is placed in the bed, he or she is evaluated by the physician. After this initial evaluation, the physician places orders for lab tests, radiology tests, specialty consults, or therapies to be conducted. Note, it is possible to deviate from this flow pattern as tests can be ordered by a nurse or before the physician sees the patient (Batt and Terwiesch, 2015). However, in our setting, this occurs less than 5% of the time, and tests are nearly always ordered by physicians. Once these are carried out and results of tests and consults are available, the physician assesses the patient's need for further care and makes a disposition decision. This can be to admit the patient to the hospital or to discharge the patient home or to an outside facility.

For this study, we leverage an exogenous change to the way RPF was disclosed to physicians at one of two EDs. We describe this change in greater detail in section 3.2. The two EDs, which we refer to as Treatment ED and Control ED, both belong to the same not-for-profit, integrated health care system in California. They are located less than 15 miles apart and serve a similar catchment area. Each has a large patient load, with more than 70,000 patient encounters each in 2010.

Though the medical centers within the health care system are interdependent and cooperative, they are each independently managed and have the ability to design their own practice structures and initiatives. The physician practice group that staffs each ED is specific to the ED; in other words, different physicians staff the two EDs. Nevertheless, Treatment ED and Control ED have four key commonalities. First, both EDs

Figure 1: Schematic Representation of the Patient Flow Process in an ED



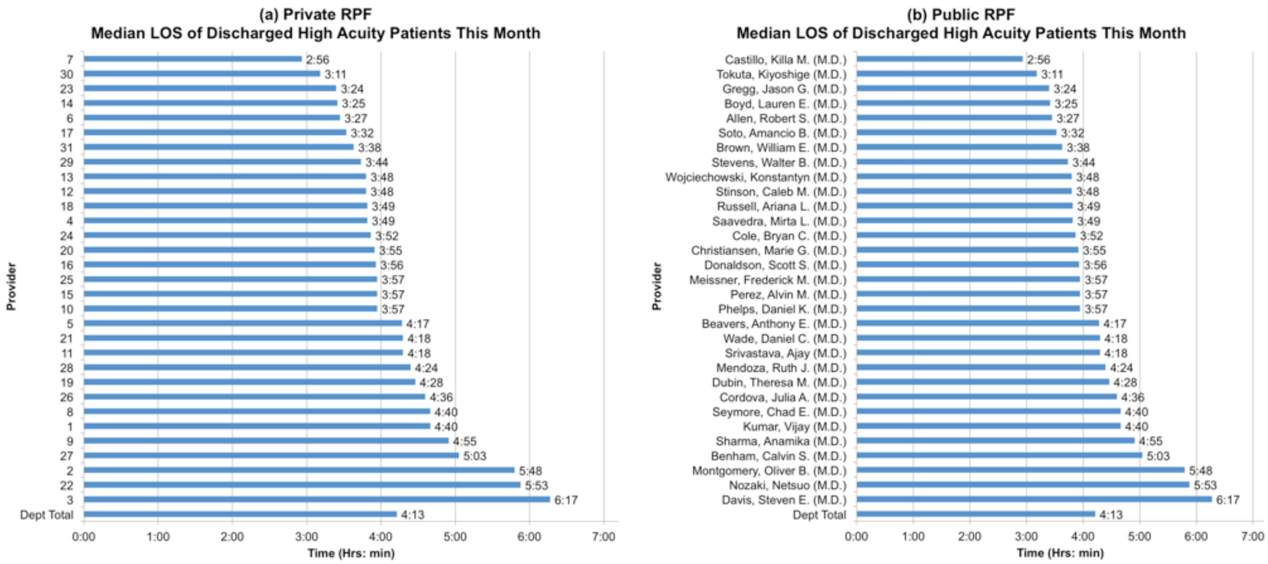
Notes. For process measures with a corresponding time stamp in the electronic health record, we note the time stamp in bold italics.

employ physicians on a salaried basis, with opportunities for additional shifts but no additional compensation for ordering more tests or working more hours on a shift than scheduled. This enables us to disentangle the effect of public RPF on productivity from the effect of a monetary incentive to outperform one's coworkers. Second, to ensure fairness among physicians in terms of their average workload, both EDs use a round robin routing policy that fairly allocates patients to physicians independent of physician work speed or idle time (Patel and Vinson, 2005). Third, prior to leaving at the end of the shift, physicians at both EDs are required to discharge, or at least complete a care plan for, each of the patients assigned to them. To make this possible, physicians have two to four hours (depending on the scheduled shift length) of protected time at the end of the scheduled shift during which they are not assigned any new patients. Fourth, both EDs have a fast track to which less urgent and non-urgent patients (ESI levels 4 and 5) are routed for their care. On a given shift, physicians work in either the main ED or the fast track.

3.2 Relative Performance Feedback at Treatment ED and Control ED

Prior to August 2010, both Treatment ED and Control ED provided private RPF to their respective group of physicians and encouraged physicians to reduce their patients' LOS. At each ED, private RPF was presented in the form of a ranked histogram of the median LOS of the patients of a given acuity level (e.g., ESI levels 1, 2, and 3) treated by each physician in that facility during a given period of time. Each physician had his or her own bar on the histogram and was identified by a code number (see Figure 2). Physicians knew their own numbers, but not the numbers for other physicians. Thus, the histogram provided each physician with information about his or her own ranking among all physicians in the ED, but did not provide any insight into which specific other physician was a top- or bottom-ranked physician.

Figure 2: Example of Monthly Report with Private versus Public RPF



Notes. Each ranked histogram reports, by physician, the median LOS of all ESI level 1 through 3 patients over a one-month period who were subsequently discharged. The examples are constructed using actual data from Treatment ED. For the public RPF report, names have been disguised using a fake name generator.

At Treatment ED, there were two LOS metrics presented to physicians in monthly reports: (a) the median LOS of all high acuity patients (ESI levels 1 – 3) who were subsequently discharged and (b) the median LOS of all low acuity patients (ESI levels 4 and 5). At Treatment ED, these metrics were presented to all physicians at the monthly staff meeting. During the "State of the ED" update portion of the meeting, Treatment ED's Chief of Emergency Medicine presented and discussed these productivity metrics. In particular, the Chief highlighted the metrics of top-ranked physicians (i.e., physicians with the shortest median LOS), although she did not identify these physicians by name. At the conclusion of the staff meeting, the ranked histogram report was posted in the staff lounge until a new report replaced it the following month. The same two LOS metrics were also presented to physicians at Control ED. Rather than at monthly meetings, these metrics were presented to each physician at Control ED twice a year at a one-on-one performance evaluation meeting with the Chief or Assistant Chief of Emergency Medicine. At both Treatment ED and Control ED, the Chiefs did not use the information to punish or shame low performers.

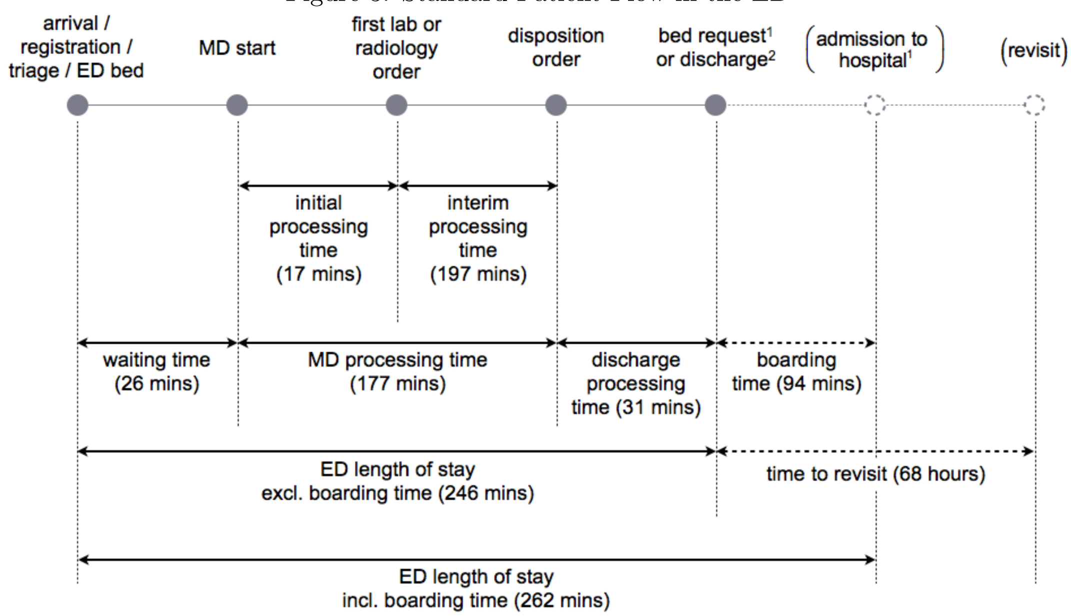
In August 2010, only Treatment ED transitioned to providing public RPF (see Figure 2). This change was implemented after Treatment ED's Chief saw an example of non-blinded productivity metrics of sales clerks being provided at a large department store. The motivation to change from private to public RPF was therefore an exogenous shock to the physicians. With public RPF, the monthly reports at Treatment ED included the same metrics as before, but now each bar in the ranked histogram was accompanied by a physician's name in lieu of a code number. Thus, with public RPF, each physician at Treatment ED could see his or her own productivity metric and how he or she performed relative to all other Treatment ED physicians, as well as identify high performers from whom he or she could learn best practices regarding workflow. During the State of the ED updates at monthly staff meetings, Treatment ED's Chief continued to highlight top-ranked physicians, now sometimes by name. In addition, top-ranked physicians—whose identities were no longer anonymous—were encouraged to share efficiency tips at the meeting, and all physicians were encouraged to shadow top-ranked physicians to learn best practices. Two main efficiency tips emerged from these meetings: (a) ordering lab and radiology tests as early in the care delivery process as possible (efficiency tip 1) and (b) beginning the discharge instructions and encounter note as soon as possible after the initial patient evaluation so that it takes less time to complete this documentation once test results are available (efficiency tip 2).

At Control ED, no changes were made to the reporting of private RPF throughout the study period. Note, throughout the study period, Control ED maintained a page on its internal website with efficiency tips for reducing patient LOS. These tips were solicited by Control ED's Chief and posted anonymously. The two efficiency tips that emerged from Treatment ED's staff meetings were among the efficiency tips listed on Control ED's internal website. Throughout the study period, no other changes were made with respect to the processes and resources in the high acuity area at only one ED and not the other.

3.3 Data

The data for our main analyses come from the electronic health records at Treatment ED and Control ED. Our data are de-identified and consist of all patient encounters of ESI levels 1, 2, and 3 from January 2009 to December 2011. The data include, but are not limited to, patient encounter-level information regarding the patient's time of arrival and departure, ESI level, attending physician, disposition, and time stamps for several process measures of the patient flow through the ED (see Figure 3). In our data, the attending physician of record is the physician who is initially assigned to the patient, even if the patient is handed off to an oncoming physician at the conclusion of the shift. Note, at both EDs, the initial physician is responsible for completing a care plan for each patient who is handed off to an oncoming physician. We exclude patients with no attending physician or ESI level listed on their record, and patients who had a LOS of less than one minute. Altogether, we exclude 1,832 of 303,014 observations or 0.6% of the overall sample. For our analyses, we also exclude data from August 2010 to account for a washout period because the exact date of the month when public RPF was made available to physicians is unknown. In addition, we limit our sample to the patients seen by physicians whose home facility was the facility to which the patient presented. Physicians who work in a facility other than their home facility tend to be those based at another ED within the health care system who are brought in to cover a portion of a shift when the facility's own physicians are not able to staff the ED (e.g., during monthly staff meetings). The resulting final sample consists of 279,025 patient encounter-level observations.

Figure 3: Standard Patient Flow in the ED



[1] For patients who were admitted to the hospital. [2] For patients who were discharged to home or to an outside facility.
Notes. Time durations (in parentheses) are mean times of high acuity patient encounters at Treatment ED and Control ED from January 2009 to December 2011. Subcomponents of time durations do not necessarily add up to the larger component because not all patients experience each process (e.g., 20% of patients do not receive a lab or radiology order, 83% of patients are not admitted to the hospital and thus do not experience boarding time). $N = 279,025$ patient encounters.

We link the electronic health record data with data from the health care system's patient satisfaction survey. This survey is designed, developed, and administered by the health care system to all patients who have been treated in one of its EDs. Surveys are distributed within 72 hours of their visit by email or regular postal mail, and the system-wide response rate in 2010 was 25%. The survey assesses patient satisfaction with regard to interactions with the physician, other health care providers (e.g., nurse, other staff), the ED overall, and ancillary services (e.g., pharmacy, laboratory, radiology). For our analyses, we focus on three survey questions that measure patients' satisfaction with their physician: (1) the physician's skills and ability, (2) the patient's confidence that the physician provided the care and services the medical condition required, and (3) how well the physician listened and explained what was being done and why. We also obtain data on two general experience measures: (1) the total time spent in the ED and (2) how well the patient's needs were met. Each of the measures are rated using a 5-point Likert scale, ranging from 1 = Poor to 5 = Excellent.

In addition, we collected data from semi-structured interviews with 41 ED physicians (21 at Treatment

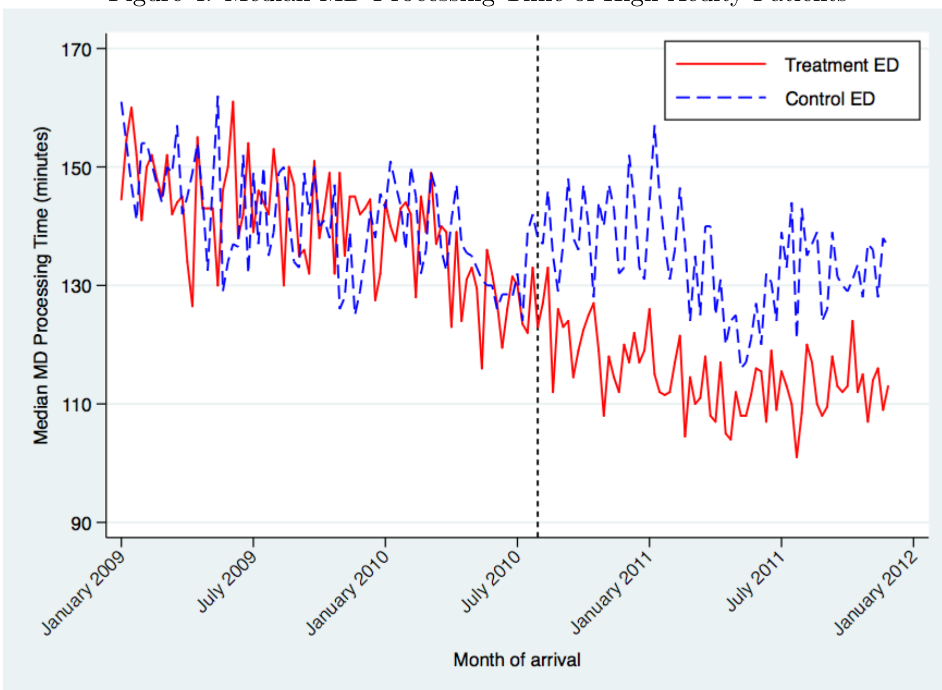
ED and 20 at Control ED) of varying levels of productivity. These interviews were conducted by the first and second authors in July 2015, at which time Treatment ED physicians were still receiving public RPF reports and Control ED physicians were still receiving private RPF reports. Interviews addressed physicians' management of workflow to reduce LOS and their experiences with the RPF system in place. In addition to several other questions, we asked each physician to list the different ways in which they manage their workflow to reduce patients' LOS without hurting quality; what information they take away from the RPF reports; how the RPF reports have affected their workflow, if at all; and whether they are motivated to be towards the top of the relative performance distribution. Each interview was recorded and transcribed. The first author reviewed all interview transcripts to identify illustrative quotes for each of the interview questions relevant to our analyses. Using a survey scale, we also asked each physician to indicate the extent to which he or she felt ashamed, deserving of criticism, stupid, self-conscious, or embarrassed when seeing the RPF reports. These terms collectively comprise an index for shame (Brown et al., 2009), and were interspersed among other terms capturing positive affects. Terms originally come from the Positive and Negative Affect Schedule (Watson et al., 1988) and the Personal Feelings Questionnaire (Harder and Lewis, 1987). Note, we are unable to link these interview data with the electronic health record data per the Institutional Review Board's requirement to de-identify all data. The first and second author also observed the workflow of two physicians each at Treatment ED and Control ED, respectively: one physician who is typically top-ranked and one physician who is typically bottom-ranked. The relative ranking of each physician was not revealed until after the observations had been completed.

4 Empirical Strategy

We rely on a difference-in-differences approach to estimate the causal effect of public RPF on physicians' productivity. By measuring the difference in differences between Treatment ED and Control ED over time, this approach allows us to control for characteristics unobservable to the researcher that may also impact performance (Imbens and Wooldridge, 2009). The implementation of public RPF constitutes our treatment, which was introduced in August 2010 as an exogenous shock to physicians at Treatment ED. Because this approach accounts for differences in baseline levels of physician productivity across the two facilities, it adjusts for any potential differences that may arise due to the difference in the frequency with which RPF reports were distributed at Treatment ED and Control ED throughout the study period.

To justify the use of this methodological approach, we first examine a key assumption: whether MD processing time exhibits parallel trends before the intervention at Treatment ED and Control ED (Abadie, 2005). We address this assumption by testing the difference in trends at Treatment ED and Control ED in the pre-intervention period using monthly time trends. The results support the assumption of parallel trends, as the difference in trends is not statistically significant at conventional levels ($p \approx 0.51$). This assumption is also visually supported by Figure 4, which shows roughly parallel trends between unadjusted MD processing times at Treatment ED and Control ED in the 19 months prior to the introduction of public RPF at Treatment ED.

Figure 4: Median MD Processing Time of High Acuity Patients



Notes. This figure depicts trends in raw data that are unadjusted for covariates, physician fixed effects, or time trends. Data are collapsed to the week level for ease of presentation. Vertical dashed line indicates the time when public RPF was implemented at Treatment ED (August 2010).

4.1 Effect of Public RPF on Physician Productivity

We begin by examining the change in physician productivity when RPF is disclosed publicly as opposed to privately. As a proxy for productivity, we measure MD processing time, which begins with the time the physician commences care and ends with the disposition order for admission or discharge (see Figure 3). We estimate the following fixed effects log-linear model at the patient encounter level:

$$\ln\left(MDProc_{ijt}\right) = \beta_0 + \beta_1 Treat_{ij} \times Post_t + \delta \mathbf{X}_{ijt} + \theta_t + \alpha_j + \epsilon_{ijt} \tag{1}$$

In equation (1), i indexes each patient encounter, j indexes each physician, and t indexes time in month-year periods. $Treat_{ij}$ equals one for patients seen by a physician at Treatment ED and zero for patients seen by a physician at Control ED. $Post_t$ equals one after public RPF was implemented (beginning September 2010) and zero before (up to July 2010). We set $Post_t$ to missing for August 2010 to account for a washout period. $\mathbf{X}_{ijt}$ is a vector of control variables that includes time-varying patient- and ED-level covariates known to impact the set of outcome measures we employ. Specifically, to adjust for heterogeneity across patient types, we control for patient age, gender, and ESI level. To account for the inverted U-shaped relationship between ED workload and worker performance (KC and Terwiesch, 2009; Tan and Netessine, 2014), we include both a linear and quadratic term for the total number of patients in the ED at the beginning of each patient encounter. To account for time-varying levels of staffing in the ED, we include the total number of physicians working on shift, with shifts operationalized as morning (7am-2pm), afternoon (3pm-10pm), and overnight (11pm-6am) shifts. In addition, we control for the time of the day and the day of the week of the patient encounter, which may have implications for the availability of resources external to the ED. θ_t accounts for time trends by controlling for the month-year of the observation. The physician fixed effect α_j allows us to control for time-invariant aspects of physicians and controls for all between-physician variance such that our model explores within-physician variance. ϵ_{ijt} are heteroskedasticity-robust standard errors clustered by physician, which addresses the serial correlation problem and relaxes the assumption that standard errors are identically distributed and independent of each other (Bertrand et al., 2004; Wooldridge, 2010). Each of the main effects for $Treat_{ij}$ and $Post_t$ are omitted due to the former being perfectly collinear with the physician fixed effects (since each physician only works at either Treatment ED or Control ED) and the latter being perfectly collinear with the month-year fixed effects (since $Post_t$ always equals zero in the month-years before August 2010 and one in the month-years after August 2010).

To test H1, our coefficient of interest is β_1. The term β_1 represents the difference in logged MD processing time between patients presenting to Treatment ED and Control ED before and after the introduction of public RPF at Treatment ED. Therefore, β_1 captures the impact of moving from a system in which RPF is disclosed privately to one in which it is disclosed publicly. We log $MDProc_{ijt}$ after adding one because the distribution of this variable is right-skewed.

4.2 Standardized Work as a Moderator of the Effect of Public RPF on Physician Productivity

To examine if the effect of public RPF on physician productivity varies by whether work is standardized, we repeat the estimation of equation (1) on two subsamples of the data: (a) patients who received a diagnosis of a heart attack (i.e., acute myocardial infarction (AMI)) or stroke, and (b) all others. We isolate AMI and stroke as the two conditions for which standardized processes are in place based on discussions with ED physicians and administrators who indicate that these were the only two conditions for which standardized protocols existed and were adhered to during the study period.

In testing H2, our coefficient of interest is β_1. We predict β_1 will not be statistically significant with the sample of patients diagnosed with AMI or stroke, and will be negative and significant with the sample of all other patients.

4.3 Effects of Public RPF on Care Intensity, Clinical Quality, and Patient Satisfaction

Given the importance of finding the optimal balance between operational efficiency and quality in service settings like the ED (Anand et al., 2010; Batt and Terwiesch, 2015; Hopp et al., 2007), we also explore whether there are any negative implications for care intensity, clinical quality, and patient satisfaction that result from publicly disclosing RPF. Using the available disposition data, we construct the following four binary indicators as proxies for care intensity: whether a lab test was ordered, whether a radiology test was ordered, whether a specialty consult was requested, and whether the patient was admitted to the hospital. Due to data limitations, we are unable to capture the number and type of lab tests, radiology tests, or specialty consults that were placed. We also use the following two binary indicators as proxies for clinical quality: whether the patient died in the ED and whether the patient returned to an ED within the health care system's network within 72 hours. Instead of limiting our definition of revisit to a patient's subsequent visit to the same facility as the initial encounter, we use a broader definition that captures a subsequent visit to any ED within the health care system's network to account for the fact that patients are not limited in their choice of which ED to visit. To assess the change in care intensity and clinical quality associated with the implementation of public RPF, we estimate the following fixed effects logit model:

$$\ln\left[\frac{\Pr\left(Q_{ijt}=1 \mid X_{ijt}\right)}{1-\Pr\left(Q_{ijt}=1 \mid X_{ijt}\right)}\right] = \rho_0 + \rho_1 Treat_{ij} \times Post_t + \delta \mathbf{X}_{ijt} + \theta_t + \alpha_j \tag{2}$$

In equation (2), Q_{ijt} equals one if patient encounter i with physician j at time t involved a lab test (or a radiology test, or a specialty consult, or a hospital admission, or an in-ED death, or an ED revisit, respectively), and zero otherwise. All other variables remain the same as in equation (1). Monte Carlo studies, like Greene (2004), have shown bias due to the incidental parameters problem is large when the number of observations per group is small, and this bias declines quickly as the number of observations per group increases. Greene's (2004) conclusions were based on simulations with the length of the panel T ranging from 2 to 20. In this study, the number of observations per physician, T_j, has an average value of 1,000. Since each physician has many observations, the bias due to the incidental parameters problem is likely to be quite small in this setting. Nevertheless, we repeat our estimation using a linear probability model and note the lack of substantive difference in our findings when using either of these specifications. To account for the nonlinearity introduced by the logit model, we obtain the average marginal treatment effect by calculating the marginal effect for each observation in the data and then averaging these results.

To assess potential changes in the level of patient satisfaction, we use data from the patient satisfaction survey described in section 3.3. Using each of the three measures of patients' satisfaction with their physician and the two general experience measures as the dependent variable, respectively, we estimate a fixed effects ordered logit model to account for the ordinal nature of the dependent variables. This ordered logit model takes a similar form as equation (2), but allows the dependent variable to take values ranging from 1 to 5.

4.4 Consideration of Potential Mechanisms

As noted in section 2.3, there are three potential mechanisms that may lead to an increase in worker productivity when switching from private to public RPF: motivation to be at the top of the relative performance distribution, motivation to avoid being at the bottom of the relative performance distribution, and the identification and diffusion of best practices that is enabled by public RPF. Empirically separating these explanations is a challenging task given they need not be mutually exclusive of one another. Acknowledging this, we explore which of these mechanisms may serve as the leading explanation, rather than trying to estimate the extent to which each independently explains the hypothesized improvement in worker productivity.

For this, we examine (a) heterogeneous treatment effects by a physician's ranking on the relative performance distribution and (b) between- and within-physician variation in initial and interim processing times, respectively. For the latter, we focus on the variation in initial and interim processing times (as opposed to total MD processing time) because each directly corresponds to the specific portion of processing time that would be affected by the adoption of efficiency tips 1 and 2, respectively (see Figure 3). Here, initial processing time begins with the time the physician commences care and ends with the first lab or radiology order. Interim processing time begins with the first lab or radiology order and ends with the disposition order for admission or discharge.

4.4.1 Heterogeneous Treatment Effects

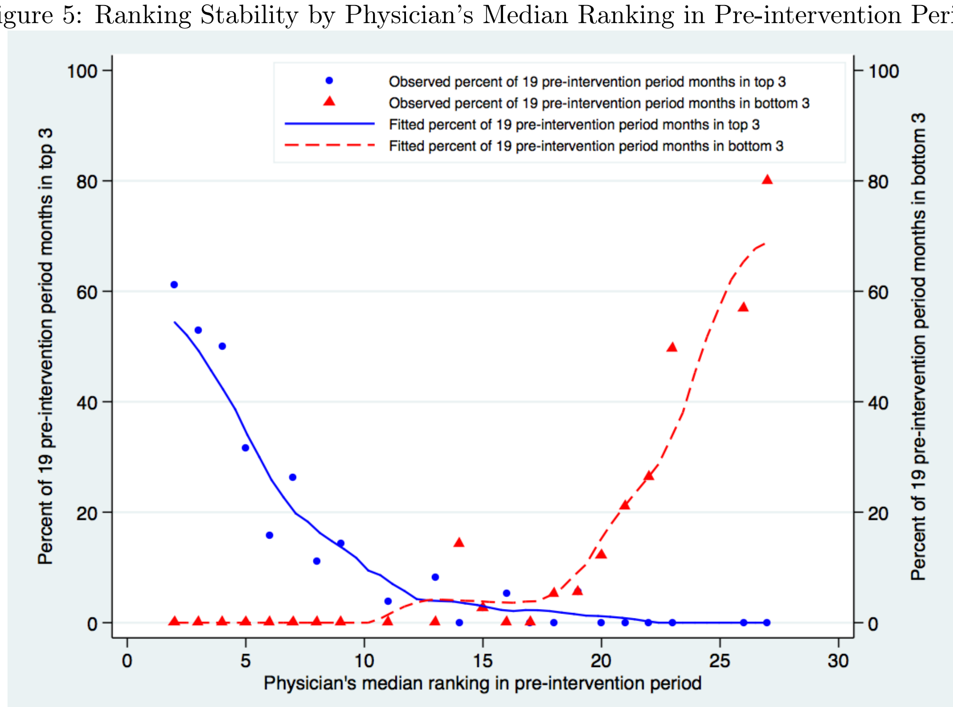
Examining heterogeneous treatment effects of public RPF on physician productivity allows us to assess whether the implementation of public RPF has a differential effect on the productivity of physicians who were initially top-ranked or bottom-ranked, respectively. If the leading explanation is the motivation to be at the top of the RPF distribution, we would see a differential increase in productivity among top-ranked physicians. In contrast, if the leading explanation is either the motivation to avoid the shame of being bottom-ranked or identifying and diffusing the best practices of top-ranked physicians, we would see a differential increase in productivity among bottom-ranked physicians as a result of public RPF implementation.

To examine the heterogeneous effects of public RPF on the productivity of top-ranked versus bottom-ranked physicians, we estimate the following fixed effects log-linear model at the patient encounter level:

$$\begin{aligned}\ln(MDProc_{ijt}) =& \lambda_0 + \lambda_1 T3_j \times Post_t + \lambda_2 B3_j \times Post_t + \lambda_3 Treat_{ij} \times Post_t \\ &+ \lambda_4 T3_j \times Treat_{ij} \times Post_t + \lambda_5 B3_j \times Treat_{ij} \times Post_t + \delta \mathbf{X}_{ijt} + \theta_t + \alpha_j + \epsilon_{ijt}\end{aligned} \tag{3}$$

In equation (3), $T3_j$ ($B3_j$) is a binary indicator for physician j that equals one for initially top-ranked (bottom-ranked) physicians and zero otherwise. We define a physician's initial level of productivity using his or her median monthly ranking in the pre-intervention period. Drawing from the ranked histograms on median LOS of all discharged high acuity patients, we use the median ranking from the 19 pre-intervention months to determine the top-ranked and bottom-ranked physicians at each facility. For our analyses, we focus on the top 3 and bottom 3 median rankings, respectively, because the rankings of physicians whose median pre-intervention rankings fall within these categories are the most stable. Figure 5 shows that physicians with a median ranking above 5 are among the top 3 physicians more than 50% of the time during the 19 pre-intervention period months and are never among the bottom 3. In addition, physicians with a median ranking of 24 or below are never among the top 3 physicians during the 19 pre-intervention months, but are among the bottom 3 in more than 50% of the 19 pre-intervention months. Thus, membership in the top 3 and bottom 3 is fairly stable.

Figure 5: Ranking Stability by Physician's Median Ranking in Pre-intervention Period



All other variables remain the same as in equation (1). The main effects for $T3_j$, $B3_j$, and $Treat_{ij}$, and the interaction effects $T3_j \times Treat_{ij}$ and $B3_j \times Treat_{ij}$ are omitted due to perfect collinearity with the physician fixed effects, and the main effect for $Post_t$ is omitted due to perfect collinearity with the month-year fixed effects. Our coefficients of interest are λ_4 and λ_5, which each represents the additional effect of being top-ranked or bottom-ranked, relative to being in neither category, on the impact on MD processing time of moving from private to public RPF.

4.4.2 Change in Between- and Within-Physician Variation in Processing Times

Examining the change in the variation in initial and interim processing times *between* physicians helps us understand how the implementation of public RPF affects the distribution of ways in which different physicians manage their workflow. If the leading explanation for improved productivity is either physicians' motivation to avoid the shame of being bottom-ranked or identifying and diffusing the best practices of top-ranked physicians, we would see a decrease in the between-physician variation in processing times. Specifically, if physicians are mimicking their high-performing peers by adopting their best practices, the set of practices used by physicians should become more homogeneous and we would expect a reduction in the between-physician variation in processing times. In contrast, if the leading explanation is the motivation to be top-ranked, we would see an increase in the level of between-physician variation.

Examining *within*-physician variation allows us to assess whether public RPF affects the distribution of ways in which each physician manages his or her workflow. If the leading explanation is that physicians are consistently adopting the best practices of their top-ranked peers, we would see a decrease in the within-physician variation of initial and interim processing times. However, if the leading explanation is either the motivation to be at the top of the distribution or to avoid being at the bottom of it and physicians are merely speeding up rather than differently managing their workflow, we would not expect to see a significant change in the within-physician variation in processing times.

To measure variation in initial and interim processing times between and within physicians, we use the coefficient of variation (CV), which is a unitless measure equal to the standard deviation divided by the mean of a random variable. For variation between physicians, we calculate the CV for initial processing time and interim processing time, respectively, for each facility-day-level observation m by dividing the standard deviation of each measure at each facility on a given day by the mean of the measure at the same unit of observation. For variation within physicians, we calculate the CV for initial processing time and interim processing time, respectively, for each physician-shift-level observation n by dividing the standard deviation of each measure for each physician on a given shift by the mean of the measure at the same unit of observation. Using these measures of variation, we assess the change in variation associated with the implementation of public RPF by estimating the following log-linear model:

$$\ln(CV_{mt}) = \varphi_0 + \varphi_1 Treat_m + \varphi_2 Treat_m \times Post_t + \delta \mathbf{X}_{mt} + \theta_t + \epsilon_{mt} \tag{4}$$

$$\ln\left(CV_{nt}\right) = \tau_0 + \tau_1 Treat_n + \tau_2 Treat_n \times Post_t + \delta \mathbf{X}_{nt} + \theta_t + \epsilon_{nt} \tag{5}$$

In equations (4) and (5), m indexes each facility-day level observation and n indexes each physician-shift level observation. Covariates are measured as the means of each covariate described in section 4.1 at the corresponding level of observation. All other variables remain the same as in equation (1). The main effect for $Post_t$ is omitted due to perfect collinearity with θ_t. Our coefficients of interest are φ_2 and τ_2, which each represent the change in logged CV between and within physicians, respectively, associated with the implementation of public RPF. We log CV to interpret these estimates as percent changes.

5 Results

Table 1 presents means for all patient- and ED-level covariates included in the empirical models stratified by facility (Treatment ED or Control ED) and time period (pre-intervention or post-intervention). At both Treatment ED and Control ED, approximately 80% of high acuity patients presenting from January 2009 to December 2011 were of ESI level 3 and only 1% of high acuity patients were of ESI level 1. The average age of patients was 45 years and 59% of patients were female. On average, there were 31 to 36 patients in the ED at a given time with 4 to 5 physicians on a given shift. Approximately 40% of shifts were morning shifts and 45% were afternoon shifts. Patients were slightly more likely to present to the ED on Mondays, Tuesdays, and Saturdays compared to other days of the week. On a given shift, each physician saw 14 patients on average and was multitasking across 4 to 5 patients at a given time (not shown in table).

Between the pre- and post-intervention periods, changes to the distribution of patient acuity and gender were very similar at both facilities, though there were small differences in the changes to the average age of patients and the total number of patients in the ED at a given time. At both sites, there was a small increase in the proportion of ESI 2 patients and a corresponding decrease in the proportion of ESI 3 patients due to an updated training on how to appropriately assign ESI levels. In addition, though the average daily patient volume increased at both facilities, this increase was more pronounced at Treatment ED. All of these differences are adjusted for in our empirical specification, and thus we account for any differential changes between Treatment ED and Control ED.

We examine the correlations between each of the control variables, none of which has correlations close to 0.80, thus minimizing concerns about collinearity. We check for multicollinearity by calculating variance inflation factors (VIF) as well. The largest VIF in our empirical model is 6.8 and the mean VIF is 2.2, both of which fall below the conventional threshold of 10 (Hair et al., 1998). This suggests that multicollinearity is not a concern (Wooldridge, 2012).

Table 1: Summary Statistics

	Treatment ED			Control ED		
Variables	Pre	Post	Diff	Pre	Post	Diff
Patient Level						
% ESI Level 1	1%	1%	0.01%	1%	1%	0.001%
% ESI Level 2	18%	23%	5.7%	19%	22%	3.1%
% ESI Level 3	82%	76%	-6.3%	81%	78%	-3.0%
Age (Years)	45	43	-1.1	46	47	0.5
% Female	59%	59%	0.1%	59%	59%	0.1%
Total no. of patients in ED	36	32	-3.3	31	32	1.2
ED Level						
Total no. of physicians on shift	4	5	0.4	4	5	0.5
% Morning shift	38%	37%	-0.7%	38%	40%	1.3%
% Afternoon shift	45%	46%	0.7%	44%	44%	-0.8%
% Overnight shift	18%	18%	0.003%	17%	17%	-0.5%
% Sunday	14%	15%	0.5%	14%	14%	-0.04%
% Monday	16%	15%	-0.3%	15%	15%	-0.02%
% Tuesday	15%	15%	0.2%	14%	14%	-0.05%
% Wednesday	13%	14%	0.3%	14%	15%	0.3%
% Thursday	14%	14%	-0.03%	14%	14%	-0.09%
% Friday	14%	14%	-0.4%	14%	14%	0.02%
% Saturday	15%	14%	-0.3%	14%	14%	-0.1%
Daily Patient Volume	131	147	16	140	145	5

Notes. $N = 279,025$ patient encounters.

5.1 Effect of Public RPF on Physician Productivity

We estimate equation (1) to examine how public RPF affects physician productivity. In the first two columns of Table 2, we estimate equation (1), first without and then with month-year fixed effects to account for time trends. We find that the estimate for the effect of public RPF on logged MD processing time remains stable at -8.6% ($p < 0.01$). This 8.6% decrease corresponds to a 17-minute decrease in MD processing time for an average patient presenting to Treatment ED. This average effect of public RPF can be seen graphically as well. In Figure 4, the median MD processing time of high acuity patients drops significantly immediately after the implementation of public RPF at Treatment ED. This supports H1, which predicts that public RPF leads to a decrease in processing time, on average.

In addition to productivity at the physician level, we also examine the effect of public RPF on system-level productivity. We find that the implementation of public RPF is associated with a significant increase in system-level productivity, as measured by ED LOS excluding boarding time (10.0% decrease, $p < 0.001$) and ED LOS including boarding time (11.3% decrease, $p < 0.001$), where boarding time is the amount of time that patients being admitted to the hospital spend waiting for an inpatient bed (Table A1 Columns 1 and 2). In addition, we find that the increase in physician productivity is not accompanied by a lower throughput rate. Following KC (2014), we measure throughput rate as the number of patients discharged by a physician

Table 2: Average and Heterogeneous Effects of Public RPF on Physician Productivity

	Average Effect					**Heterogeneous Effect**
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Full Sample	Full Sample	AMI or Stroke	Not AMI or Stroke	Abdominal Pain	Full Sample
Post X Treat	-0.086**	-0.086**	0.037	-0.086**	-0.124***	-0.080**
	(0.027)	(0.028)	(0.061)	(0.028)	(0.030)	(0.030)
*Post X Treat X T*3						-0.068
						(0.074)
*Post X Treat X B*3						-0.085**
						(0.030)
ESI level 2	0.269***	0.271***	0.076	0.276***	-0.069	0.270***
	(0.030)	(0.029)	(0.088)	(0.030)	(0.130)	(0.033)
ESI level 3	-0.120***	-0.122***	0.245**	-0.120***	-0.279*	-0.124***
	(0.033)	(0.033)	(0.088)	(0.033)	(0.131)	(0.037)
Age	0.011***	0.011***	-0.000	0.011***	0.013***	0.011***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Female	0.048***	0.048***	-0.002	0.049***	0.068***	0.047***
	(0.005)	(0.005)	(0.031)	(0.005)	(0.012)	(0.005)
Total patients	0.009***	0.006*	0.006	0.006*	0.003	0.005
	(0.003)	(0.002)	(0.011)	(0.002)	(0.005)	(0.002)
$(Total\ patients)^2$	-0.001***	-0.000**	-0.001	-0.000**	-0.000	-0.000**
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
MDs on shift	-0.010***	-0.005*	-0.014	-0.005*	-0.000	-0.005*
	(0.002)	(0.002)	(0.013)	(0.002)	(0.004)	(0.002)
Afternoon shift	-0.032***	-0.032***	0.039	-0.032***	-0.063***	-0.035***
	(0.008)	(0.007)	(0.034)	(0.007)	(0.012)	(0.008)
Overnight shift	-0.022*	-0.015	0.044	-0.014	-0.001	-0.018
	(0.010)	(0.010)	(0.053)	(0.010)	(0.016)	(0.011)
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	261,586	261,586	1,930	259,656	35,084	232,297
Adjusted R-squared	0.188	0.190	0.046	0.190	0.203	0.181

Notes. Regressions are fixed effects log-linear difference-in-differences models estimated at the patient encounter level. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p <0.05, **p <0.01, ***p <0.001.

on a given shift, adjusted for the duration of the shift. We find a modest increase in throughput rate (2.6%, $p < 0.05$), which suggests that physicians are not working less to compensate for being more productive (Table A1 Column 3).

5.2 Standardized Work as a Moderator of the Effect of Public RPF on Physician Productivity

In Columns 3 and 4 of Table 2, we repeat the estimation of equation (1) on a subsample of patients with conditions with standardized processes (i.e., patients presenting with symptoms of an AMI or stroke) and those with conditions without standardized processes (i.e., all other patients). We find that patients presenting with symptoms of an AMI or stroke experience no significant change in MD processing time after the implementation of public RPF ($p \approx 0.54$). In contrast, patients who do not fall into this category exhibit an 8.6% decrease in MD processing time with public RPF ($p < 0.01$). As a test of robustness, we also consider a third subsample that is limited to patients presenting with abdominal pain symptoms—an example cited by physicians as a frequent and potentially serious case for which standardized processes are not in place (Column 5). We find that these patients experience a 12.4% decrease in MD processing time with public RPF ($p < 0.001$). This strongly supports H2, which predicts that the decrease in processing time resulting from a shift from private to public RPF is greater when standardized processes are not in place compared to when they are.

5.3 Effect of Public RPF on Care Intensity, Clinical Quality, and Patient Satisfaction

Given the potential tradeoff between operational efficiency and service quality in service settings like the ED (Anand et al., 2010; Hopp et al., 2007; Netessine and Yakubovich, 2012), we consider the impact of public RPF on care intensity, clinical quality, and patient satisfaction. We estimate equation (2) to explore these effects and present results in Table 3. Because the binary-outcome logit regression model is nonlinear, we express results as marginal effects averaging across all values of the covariates in the data.

Table 3: Marginal Treatment Effect of Public RPF on Care Intensity and Clinical Quality

	Care Intensity				**Clinical Quality**	
Variables	(1) Lab Ordered	(2) Radiology Ordered	(3) Consult Requested	(4) Admitted	(5) Expired	(6) Revisit 72 hours
Post X Treat	-0.029***	-0.038***	0.001	0.012**	-0.001*	-0.004***
	(0.009)	(0.007)	(0.006)	(0.005)	(0.000)	(0.000)
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277,970	278,473	277,945	278,039	264,876	271,956
Pseudo R-squared	0.106	0.103	0.140	0.154	0.553	0.033

Notes. Regressions are fixed effects logit difference-in-differences models estimated at the patient encounter level and coefficients are expressed as marginal effects. Controls not shown include ESI level, age, gender, current patient count, current patient count squared, total number of MDs on shift, arrival shift type, time of day, day of week, month-year fixed effects, and physician fixed effects. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p <0.05, **p <0.01, ***p <0.001.

In Columns 1 – 4, we examine the change in care intensity under public RPF by estimating the difference in the change in a patient's likelihood of having a lab test ordered, a radiology test ordered, a specialty consult requested, or being admitted to the hospital, respectively, at Treatment ED and Control ED. With regard to care intensity *in* the ED, we find that the implementation of public RPF is associated with a 2.9 percentage point differential decrease in the likelihood of having a lab test ordered ($p < 0.001$), a 3.8 percentage point differential decrease in the likelihood of having a radiology test ordered ($p < 0.001$), and no significant change in the likelihood of having a specialty consult requested ($p \approx 0.90$). These results show that public RPF may be associated with a reduction in the intensity of care provided in the ED. Given top-ranked physicians were less likely to exhibit high levels of care intensity in the pre-intervention period, this decrease in in-ED care intensity may be suggestive of a convergence around best practices. When we examine changes in care intensity *beyond* the ED, we find that patients presenting to Treatment ED experience a 1.2 percentage point differential increase in the likelihood of being admitted to the hospital ($p < 0.01$). This suggests that there may be some degree of task shifting occurring, where ED physicians may be reducing MD processing time by, on the margin, deciding to admit a patient rather than keeping the patient in the ED for further observation.

Using a mediation model (Preacher and Hayes, 2008), we examine the extent to which the decrease in in-ED care intensity and the increase in task shifting explain the observed decrease in MD processing time. When we account for the four proxy measures as potential mediators, we find a statistically significant, but smaller decrease in processing time (-4.4%, $p < 0.05$), which is suggestive of partial mediation. That is, the change in care intensity contributes to the decrease in processing time, but does not fully explain the effect.

Without any additional information, it is difficult to determine whether this reduction in in-ED care intensity and some degree of task shifting is suggestive of lower quality care. If it were, we might find a corresponding increase in the likelihood of dying in the ED or returning to the ED. As shown in Columns 5 and 6, we find that public RPF is associated with a 0.1 percentage point differential *decrease* in the likelihood of dying in the ED and a 0.4 percentage point differential *decrease* in the likelihood of returning to the ED. Though economically modest, both of these decreases are statistically significant at conventional levels ($p < 0.05$ and $p < 0.001$, respectively). These findings help reduce concern that public RPF may be associated with a decrease in clinical quality.

Table 4: Marginal Treatment Effect of Public RPF on Patient Satisfaction

Variables	**Experience with Physician**			**General Experience**	
	(1)	(2)	(3)	(4)	(5)
	MD Skills and Ability	MD Provided Proper Care	MD Listened and Explained	Time Spent in ED	Needs Met
Post X Treat	-0.002	-0.002	-0.004	-0.002	-0.001
	(0.002)	(0.003)	(0.003)	(0.005)	(0.004)
Time-varying controls	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes
Observations	70,274	73,107	72,963	75,578	77,005
Pseudo R-squared	0.015	0.016	0.016	0.013	0.010

Notes. Regressions are fixed effects logit difference-in-differences models estimated at the patient encounter level and coefficients are expressed as marginal effects. Controls not shown include ESI level, age, gender, current patient count, current patient count squared, total number of MDs on shift, arrival shift type, time of day, day of week, month-year fixed effects, and physician fixed effects. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p <0.05, **p <0.01, ***p <0.001.

It is also possible that reductions in MD processing time may affect patient satisfaction, concerning both their satisfaction with the physician and their general ED experience. On the one hand, patients may feel rushed by the physician and perceive that the physician is not taking enough time to listen to them or explain what is being done and why. On the other hand, being able to leave the ED sooner may lead patients to feel more satisfied. Table 4 reports the marginal effects resulting from estimating ordered logit models of patient satisfaction. In Columns $1-3$, we find no significant change in patient satisfaction regarding their experience with the physician. In Columns 4 and 5, we find no significant change in patient satisfaction regarding their general ED experience ($p > 0.24$ for each measure).

5.4 Consideration of Potential Mechanisms

5.4.1 Heterogeneous Treatment Effects

In Column 6 of Table 2, we report results from estimating equation (3). We find that the effect of public RPF on MD processing time is significantly greater for physicians who ranked in the bottom 3 in the pre-intervention period than for mid-ranked physicians (i.e., physicians not in the top or bottom 3). In addition, we find that the differential effect for physicians who ranked in the top 3 is not statistically significant ($p \approx 0.63$). Specifically, bottom-ranked physicians attained an additional 8.5% decrease in MD processing time ($p < 0.01$) above and beyond the 8.0% decrease ($p < 0.01$) attained by mid-ranked physicians. In terms of effect size, this corresponds to an additional time savings of 20 minutes on average by bottom-ranked physicians at Treatment ED, above and beyond the 15-minute decrease in MD processing time attained by mid-ranked physicians at the same facility. Together, this points to a 35-minute time savings attained by bottom-ranked physicians at Treatment ED after the implementation of public RPF.

5.4.2 Change in Between- and Within-Physician Variation in Processing Times

In Table 5, we report results from estimating equations (4) and (5). In Columns 1 and 2, we see that the implementation of public RPF leads to a decrease in between-physician variation in both initial and interim processing times. The between-physician CV of initial processing times decreases by 14.3% ($p < 0.001$) and the between-physician CV of interim processing times decreases by 8.6% ($p < 0.001$). This suggests that different physicians may be converging on a similar set of practices with regards to how they structure their workflow. In Columns 3 and 4, we see that the within-physician CV of initial processing time decreases by 7% ($p < 0.001$) whereas the within-physician CV of interim processing time increases by 1% ($p < 0.001$). This suggests that each physician may be taking a more consistent approach with regards to how he or she structures workflow during the initial processing time, but that there is a modest increase in the variation with which each physician structures his or her workflow during the interim processing time. In additional analyses that examine heterogeneous effects in the change in within-physician CV (not shown in table), we find that the decrease in within-physician CV in initial processing times is greater for bottom-ranked physicians than for top-ranked physicians ($p < 0.01$). For the change in within-physician CV in interim processing times, this difference is not statistically significant ($p \approx 0.72$). Nevertheless, on balance, we find that the within-physician CV of total MD processing time is reduced (-2.8%, $p < 0.001$).

Table 5: Effect of Public RPF on Between- and Within-Physician Variation in Processing Times

	Between Physicians		**Within Physician**	
Variables	(1) Logged CV of Initial Processing Time	(2) Logged CV of Interim Processing Time	(3) Logged CV of Initial Processing Time	(4) Logged CV of Interim Processing Time
Post X Treat	-0.143***	-0.086***	-0.070***	0.010***
	(0.010)	(0.010)	(0.006)	(0.002)
ESI level 2	0.288	0.229	-0.026	-0.001
	(0.290)	(0.281)	(0.025)	(0.011)
ESI level 3	0.240	0.166	-0.040	0.005
	(0.286)	(0.278)	(0.027)	(0.011)
Age	0.001	0.002	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)
Female	-0.009	0.029	0.003	-0.001
	(0.054)	(0.053)	(0.006)	(0.002)
Total patients	0.000	0.051***	-0.003	0.003**
	(0.007)	(0.006)	(0.002)	(0.001)
$(Total\ patients)^2$	0.000	-0.002***	0.000***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
MDs on shift	-0.009*	-0.003	-0.001	-0.003***
	(0.004)	(0.004)	(0.001)	(0.001)
Afternoon shift	0.064	0.026	-0.003	-0.002
	(0.048)	(0.047)	(0.005)	(0.002)
Overnight shift	-0.009	0.176**	-0.011	-0.006
	(0.067)	(0.065)	(0.007)	(0.003)
Time-varying controls	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes
Physician FE	No	No	Yes	Yes
Observations	2,128	2,128	27,049	27,049
Adjusted R-squared	0.637	0.591	0.579	0.951

Notes. Regressions are fixed effects log-linear difference-in-differences models estimated at the facility-day level. Standard errors (in parentheses) are heteroskedasticity robust. *p <0.05, **p <0.01, ***p <0.001.

5.4.3 Additional Considerations Using Qualitative Data

Using data from the physician interviews, we further consider each of the potential mechanisms for improved productivity. Though we are unable to test changes in interview responses over time as we do not have data from the pre-intervention period, we are able to compare cross-sectional differences in responses across the two facilities. Note, the physicians at Treatment ED and Control ED are employed by the same health care system, work in the same city, and are under the same senior ED leadership that exists at the level of the health care system. Consequently, physicians from the two EDs are subject to an overlapping set of organizational and contextual factors that serve to increase their comparability. We summarize the findings from these interviews in Table 6.

At Treatment ED, 52% of respondents reported feeling motivated to be top-ranked on the RPF reports in comparison to 53% of respondents at Control ED. Using a two-tailed t-test assuming unequal variance, we find that these proportions are not significantly different at conventional levels ($p \approx 0.99$). Respondents at both EDs also reported similar levels of shame as measured by the shame index. On a scale ranging from 1 (Low) to 5 (High), respondents at Treatment ED reported a mean level of shame of 1.52 and those at Control ED reported a mean level of 1.34. This difference is not statistically significant ($p \approx 0.18$). Since we would expect feelings of shame to be most salient among bottom-ranked physicians, we also compare the mean level of shame reported by physicians at Treatment ED and Control ED who are ranked in the bottom third; we find no statistically significant differences between their responses ($p \approx 0.47$). However, *within* each facility, we find that bottom-ranked physicians reported a higher level of shame than the top-ranked physicians, as we would expect. Collectively, these results do not support the notion that public RPF improves productivity by increasing the level of shame felt by the bottom-ranked physicians when compared to private RPF.

When asked to list the different ways in which he or she manages workflow to reduce patients' LOS, we find significant differences between respondents at Treatment ED and Control ED in their rates of employing workflow practices that correspond to efficiency tips 1 and 2. While 100% of respondents at Treatment ED reported placing lab and radiology orders as early as possible, only 60% of respondents at Control ED reported this to be the case ($p < 0.001$). Regarding early initiation of the discharge instructions and the encounter note, 60% of respondents at Treatment ED and 30% of respondents at Control ED reported employing this practice ($p < 0.001$). This is consistent with what we would expect to find if workers receiving public RPF were to identify and adopt the best practices around workflow that are employed by their top-ranked peers.

During the interviews, several physicians at Treatment ED shared examples of how they were able to learn from the best practices of their top-ranked peers once public RPF enabled them to identify these individuals (see illustrative quotes in Table 6). Specifically with regards to efficiency tip 1, several Treatment ED physicians stated that they had changed their practices such that they no longer "dribble ordered" tests (i.e., ordered tests serially in which the physician waited for the results of the first test prior to ordering a subsequent test), and instead ordered them simultaneously so that they could be processed in parallel. Though in some cases this may have resulted in inefficiencies from ordering unnecessary tests, several physicians suggested that this was unlikely to be the case because now they were more selective about whether to order a given test at all. Whereas before physicians would often order a test to rule out a diagnosis, now they tended to do this less often by not ordering tests that were only for that reason. In this setting, physicians had the confidence to alter their workflow in this way because the integrated health care system made it possible for patients to

Table 6: Summary of Data from Physician Interviews

	Motivation to be top-ranked (1=Yes, 0=No)	**Shame Index (1=Low, 5=High)**	**Efficiency tip 1** (n (%))	**Efficiency tip 2** (n (%))	**Favor public RPF** (n (%))	**Illustrative quotes about public and private RPF reports**
Treatment (N=21)	0.52	1.52	21 (100%)	21 (60%)	17 (82%)	"The open data [was] a factor in me identifying [Dr. X]... [Before when] a patient would come in with flank pain and I would think 'maybe it's a kidney stone, maybe not,' I would order a urinalysis. If the urinalysis shows blood in the urine, then I would do a CT scan. What would happen inevitably is that it would take 2 hours to get the urinalysis results. And then sure enough, there is blood in the urine, [and] then I would order the CT scan. So that would delay the patient. [Now], based on my exam and the history, [if] I think this patient has a kidney stone, I order everything for the kidney stone and I don't wait to determine if there's blood in the urine." "We ask people at the top who are successful and have them share their ideas." "I really do feel that it is so much more helpful to have it [public]... There is so much more data to be gained and understood as you look at the group and not just look at the number and have no idea what that means." "There is value for other EDs to follow suit. I think [public] reports are fun." "With [private RPF], it wasn't as helpful. Really all I could take away was where I was in comparison to the group. I definitely get more out of the data now than I did back then."
Control (N=20)	0.53	1.34	12 (60%)	6 (30%)	18 (33%)	"With these [private] reports, I look at where I am versus everyone else. But I don't know what to do with it." "I don't know how I'd know how I could improve, like what are the limitations, what's causing that. I don't know what I [can] really take away from [private reports] other than 'Okay, that's my number. Let's try and improve it." "With [public RPF], I think it would be interesting to see which docs fell where in terms of assessing a pattern." "[With public RPF], I think I might look at who was doing 'better' or had a better metric and try to figure out what they were doing that I wasn't. It might introduce some healthy competition into things." "I think for the most part, our docs are working as hard as they can, and to have this additional pressure put on them, I don't know if it's the best thing to do in terms of morale."
p-value	0.99	0.18	< 0.001	< 0.001	< 0.01	

easily follow-up with an outpatient department if needed. These statements suggest that a patient's likelihood of receiving at least one test may have *decreased* after the intervention, which is consistent with our findings in Columns 1 and 2 of Table 3.

Also during these interviews, several Treatment ED physicians stated that knowing the identities of the physicians corresponding to various levels of performance improved the credibility of the RPF reports and the credibility of the efficiency tips, which ultimately motivated them to change their workflow. At Treatment ED, knowing the identities of top-ranked physicians via public RPF allowed physicians to assess whether these physicians provided high quality care (based on their own observations). When this was the case, other physicians were more willing to seek out and adopt their workflow practices. In contrast, at Control ED, several physicians reported skepticism around the value of the private RPF reports. Furthermore, despite there being an internal website with anonymized efficiency tips, none of the Control ED physicians we interviewed reported accessing these tips. This suggests that knowing the identities of the individuals sharing efficiency tips is an important factor in motivating peers to adopt best practices.

Finally, despite concerns raised by prior literature and critics of public RPF about its potential negative effects (e.g., demoralization, internal competition), the majority of respondents at Treatment ED (82%) favored keeping public RPF reports rather than reverting to private RPF reports. Interestingly, only 33% of respondents at Control ED reported that they would favor moving from private to public RPF reports, suggesting a discomfort with the *concept* of public RPF as opposed to the actual experience of it. In Table 6, we include illustrative quotes that capture respondents' thoughts in response to this question as well.

5.5 Robustness Checks and Specification Tests

We conduct a number of additional analyses to examine the robustness of our main findings. We can group our additional analyses into three main categories: tests of validity concerning our difference-in-differences approach, tests using alternative measures, and tests with sample exclusions. We conduct these analyses on all dependent variables, but focus our discussion here on results concerning physician productivity for brevity. Findings of additional analyses with the other dependent variables are not meaningfully different from those concerning physician productivity.

To test the internal validity concerns that arise from employing a difference-in-differences approach, we begin by examining whether MD processing time was already differentially decreasing at Treatment ED before public RPF was implemented. Following Autor (2003), we use a leads and lags model to explore the presence of pre-intervention time trends in our data. Specifically, we create five pre-period (i.e., lead) indicator variables corresponding to each of the 5 quarters before public RPF was implemented and 5 post-period (i.e., lag) indicator variables corresponding to each of the 5 quarters after implementation. We estimate a model similar to equation (1) that replaces *Post* with the ten lead and lag indicator variables. In this model, the reference category is the quarter during which public RPF was implemented. As we see in Figure A1, none of the coefficients on the five lead indicator variables is statistically significant. This suggests that Treatment ED did not exhibit a differential decrease in MD processing times in advance of the implementation of public RPF.

As another test of validity, we conduct placebo tests to assess whether our findings are merely artifacts of the structure of our data (Bertrand et al., 2004). Following the methodology employed by Pierce et al. (2015) and Staats et al. (2015), we randomly assign an implementation date and repeat the estimation of equation (1) by replacing the actual implementation date with the placebo date. We repeat this 100 times and present the results in Figure A2. We find that only three of the 100 placebo models produces a coefficient that is statistically significant at the 5% level. Each of these coefficients is weaker in statistical significance compared to that of our true estimate.

As a third test of validity, we repeat our estimation of equation (1) while limiting our study period to nullify the possibility of other changes having differentially impacted the two EDs. Specifically, we limit the study period to span a period of 3 months, 6 months, and 12 months, respectively, before and after the intervention. Compared to our main findings, we find that the effect of public RPF implementation on physician productivity is weaker with a 3-month pre/post period, suggesting a ramp-up period, but then is stronger and fairly constant from 6 months pre/post onward, reducing concern that our results are driven by the time window used for the analysis (Table A2 Columns 1 − 3).

Next, we consider several alternate model specifications. First, instead of accounting for a washout period by excluding data from August 2010, we do not impose this exclusion on the data (Table A2 Column 4). We obtain a similar set of findings, in which the implementation of public RPF is associated with an 8.3% decrease in MD processing time ($p < 0.01$).

Second, instead of using log-transformed dependent variables in estimating equation (1), we use linear dependent variables measured in minutes. The results are robust to our main findings. In particular, we find that the implementation of public RPF leads to a 17-minute decrease in MD processing time (Table A2 Column 5), which closely corresponds to the effect size estimate discussed in section 5.1.

Third, we use an interrupted time series model rather than a difference-in-differences model to estimate the effect of public RPF on MD processing time. An interrupted time series model uses data that are collected at multiple time points before and after an intervention. Unlike a difference-in-differences model, which assumes parallel trends in the treatment and control groups before the intervention and only accounts for a change in the level (i.e., mean) of the treatment group as compared to the control group, an interrupted time series model can account for both a change in level and a change in trends (Shadish et al., 2002; Wagner et al., 2002). Using an interrupted time series model, we find a change in levels but not a change in trends as a result of public RPF at Treatment ED (Table A2 Column 6). Specifically, we find an 8.8% decrease in the level of MD processing time ($p < 0.001$) and a non-significant change in trends ($p \approx 0.4$). Thus, our results are robust to whether we employ a difference-in-differences model or an interrupted time series model.

Fourth, we employ a set of alternative thresholds for determining top-, mid-, and bottom-ranked physicians. Our results concerning the heterogeneous effects of public RPF on physician productivity are robust to a more restrictive threshold of top 2 and bottom 2. With this alternate threshold, each of the coefficients of interest is similar in direction and magnitude as in our main specification (Table A3). As we adopt a progressively less restrictive threshold of top 4 and bottom 4, and top 5 and bottom 5, respectively, the average effect embodied by the mid-performing physicians remains relatively robust but the additional effect of the low-performing physicians decreases in magnitude and statistical significance. This result is consistent with the fact that physician rankings are very stable in the extremes (top 3 and bottom 3) but less so beyond this threshold, as seen in Figure 5.

Lastly, we apply various exclusions to our study sample to create a more homogeneous patient population and to better control for factors outside a physician's control. Specifically, we exclude patients with a LOS excluding boarding time greater than 48 hours ($n = 206$) and patients who received a psychiatry consult ($n = 11,312$) because the extended MD processing times of most of these patients are typically driven by placement logistics rather than by a physician's level of productivity. We exclude patients of ESI level 1 (i.e., patients needing resuscitation) ($n = 2,213$) and patients who died in the ED ($n = 669$) because their MD processing times are likely to be driven by factors other than physician productivity. Lastly, we exclude patients who left without being seen ($n = 3,185$), patients who left against medical advice, and patients who eloped because their LOSs are likely to be truncated as a result of their departure. Our results are robust to each of these exclusions (Table A4).

6 Discussion and Conclusions

In this paper, we identify public RPF as an effective tool for improving productivity by fostering standardized workflow in complex service systems. In our field setting, public RPF was disclosed to physicians in a way that it (a) provided them with specific information on their levels of productivity and (b) helped to improve their understanding of how to improve their productivity while (c) accounting for natural productivity fluctuations over time by reporting data at the month level—three key considerations identified by Netessine and Yakubovich (2012) for a successful application of RPF systems. Using patient encounter data from hospital EDs, we find that public RPF is associated with a significant improvement in productivity and no significant reduction in overall quality. We find that public RPF is particularly helpful in improving productivity when standardized work is not in place. Additional analyses suggest that these improvements are most likely driven by the identification and diffusion of best practices that is enabled by public RPF.

Specifically, we find that public RPF is associated with an 8.6% decrease in MD processing time on average. When stratified by a physician's median ranking in the pre-intervention period, we find that top- and mid-ranked physicians exhibit an 8.0% decrease in MD processing time and bottom-ranked physicians exhibit an additional 8.5% decrease in MD processing time. Though we do not find evidence of lower levels of clinical quality or patient satisfaction associated with the implementation of public RPF, we do find a decrease in in-ED care intensity as measured by the likelihood of ordering at least one lab test or radiology test. Unfortunately, we are unable to examine changes in the *number* of tests ordered as a more precise measure of care intensity, as we do not have access to this information. However, interviews with physicians suggest that the decrease in the likelihood of ordering at least one test stems from physicians acting with greater confidence on their diagnoses and being more selective about whether to order tests when they may only serve to rule out diagnoses.

6.1 Theoretical Contributions

This paper makes several contributions to the literatures on feedback, best practice transfer, and workflow management. First, our paper answers the call for field research on the impact of public—as opposed to private—feedback (Blanes i Vidal and Nossol, 2011). We do this by examining the effect of publicly disclosing RPF, a type of feedback, to ED physicians internal to an organization over a period of 3 years. We find that the small and financially costless change of replacing code numbers with worker names can lead to a substantial improvement in productivity. This study illustrates an example of how behavioral levers can be used to improve productivity by changing the way in which discretionary workers manage their work.

Second, in our field setting, we find that public RPF that is provided at a monthly interval is sufficient to improve worker productivity. This is in contrast to the experiments in Bendoly (2013) and Schultz et al. (1999), which provide continual performance feedback during task execution. Establishing the optimal frequency of feedback is beyond the scope of this paper and left for future research. On the one hand, shorter time frames between feedback reports may increase adherence to best practices. However, this may also increase workers' levels of stress (Bendoly, 2013; Schultz et al., 1999) and increase the level of spurious variation in the rankings, thus reducing their usefulness (Netessine and Yakubovich, 2012).

Third, we find that public RPF is most helpful when the productivity metrics on which the relative rankings are based are ones that can be improved by the spread of best practices rather than reflective of differences in individual ability. In our study, we find that the productivity gains from feedback are much greater for the slowest workers than for the fastest workers. This is similar to Schultz et al. (1999) but in contrast to Ashraf et al. (2014) and Barankay (2012), which find that bottom-ranked workers become discouraged and expend less effort as a result of receiving feedback about their poor performance. The key difference between our study and these latter two studies is that in the latter, bottom-ranked workers are not equipped with efficiency tips on how to improve as they are in our study. Thus, if public RPF is to be leveraged to foster improvements in productivity, the metrics on which the relative rankings are based must be carefully selected to be ones that reflect the extent to which best practices are employed.

Finally, with regard to workflow management, our study suggests that when workers have discretion in how to carry out their work and prioritize among different tasks (Hopp et al., 2009), focusing on improving the management of workflow may be particularly useful. We build on Dobson et al.'s (2013) study of the importance of workflow rules around which patient to serve (e.g., to bring in a new patient or to discharge an existing patient) to show that workflow decisions within the course of caring for a given patient (e.g., ordering all tests at once rather than serially) can also have a significant impact on worker productivity.

6.2 Implications for Practice

Our study suggests ways in which organizations may improve performance by leveraging existing efforts around performance feedback. We propose that public RPF can equip workers to learn from their peers how to better manage their workflow, particularly when work tasks are not standardized. Because this type of best practice sharing is most likely to occur internally within an organization or work group as opposed to across organizational boundaries, future research could examine whether public RPF—in which relative performance data are made available to workers within an organization or work group—is more effective than public reporting—in which performance data are made available to the external public—at spreading best practices and improving performance.

Managers currently employing public RPF may be able to further leverage its productivity benefits by segmenting the relative performance distributions to reflect worker performance when carrying out standardized versus unstandardized work tasks. This may facilitate the identification of best practices for carrying out unstandardized work tasks by highlighting workers who do particularly well with those types of tasks.

In the ED setting in particular, we find significant productivity benefits associated with implementing public RPF. We find that each high acuity patient seen by an average physician at Treatment ED experiences a time savings of 17 minutes on average in terms of their MD processing time. For those seen by an initially bottom-ranked physician at Treatment ED, the corresponding time savings is 35 minutes on average. With approximately 140 high acuity patients presenting to the ED each day, this is roughly equivalent to an average time savings of 40 patient-hours per day. With 3 hours being the average MD processing time for each patient, this suggests that Treatment ED could see an additional 13 patients per day without investing in any additional resources. Given the significant need for efficiency gains in EDs across the country, these time savings and their associated cost implications have significant practical implications.

6.3 Limitations and Future Research

This study has limitations and its results should be interpreted accordingly. First, in this study, we cannot separately identify the effect of implementing public RPF from the effect of sharing top performers' efficiency tips. This is because efficiency tips were systematically shared at Treatment ED only once public RPF enabled the identification of top performers, and not when RPF was disclosed privately. To assess whether these effects are complementary or independent, future work could identify the effects of the following four conditions in a laboratory setting: no efficiency tips with private RPF, no efficiency tips with public RPF, efficiency tips with private RPF, and efficiency tips with public RPF. Furthermore, future work could examine whether the credibility instilled in the efficiency tips by attributing them to high performers mediates the effect of public RPF on improved productivity, as is suggested by the interviews.

Second, because this study was conducted in the field, it is difficult to ascertain that the identities of coworkers remained completely anonymous under the private RPF condition. Because the code number associated with each physician typically did not change from month to month, it is possible for physicians to have had a reasonable guess regarding the identities of the top-performing physicians and to have asked these physicians for efficiency tips. Though these discussions did not happen in a structured manner during staff meetings when RPF was disclosed privately, they may have happened informally outside of these meetings given the improvement culture that existed at these EDs. Nevertheless, if this were the case, then our findings would be an underestimate of the actual effects associated with public RPF.

Third, due to restrictions on data access, we are limited in our ability to conclusively explain how the intervention affected the total number and type of tests ordered by physicians at Treatment ED. Because a system that is separate from the main electronic health record system captures the data on the number and type of tests ordered, and because we would need patient identifiers to link these two datasets, we were unable to access these data for this study. We leave further exploration of changes in test ordering behaviors as an area for future research.

Fourth, the generalizability of our findings with regard to the changes in care intensity may be limited due to the specifics of our field setting. Because the health system in which these EDs belong is a managed care entity, which is one that seeks to reduce unnecessary health care costs, a reduction in care intensity has *positive* revenue implications as opposed to negative revenue implications that would be typical in a fee-for-service system. Future work should examine how public RPF and other levers to improve productivity may affect the intensity of care in fee-for-service systems, in which services are paid for separately and therefore providers are incentivized to increase their care intensity.

6.4 Conclusions

In recent years, more organizations have been employing public RPF (Netessine and Yakubovich, 2012). There are examples of such implementations not only in health care (Gorman, 2015) but also in other service industries, such as the restaurant industry (Vanek Smith, 2015). A deeper understanding of the conditions under which public RPF leads to improved operational performance is essential to determining when and how to adopt these practices in the hopes of improving productivity. Our paper provides initial evidence suggesting that public RPF may lead to improved productivity by enabling low-performing workers to better manage their workflow by adopting the best practices of their high-performing peers. This is essential to improving the productivity of many organizations in which it is difficult to standardize each specific work task. Furthermore, being able to improve the performance of the bottom tier of employees is an important objective in many industries, such as health care and education, where poor performance of any employee can have serious repercussions for their customers. We look to future work to identify other ways in which organizations can attain these goals without sacrificing quality.

References

Abadie, A. 2005. Semiparametric Difference-in-Differences Estimators. *Review of Economic Studies* **72**(1) 1–19.

Anand, G., J. Gray, E. Siemsen. 2012. Decay, Shock, and Renewal: Operational Routines and Process Entropy in the Pharmaceutical Industry. *Organization Science* **23**(6) 1700–1716.

Anand, K. S., M. F. Pac, S. K. Veeraraghavan. 2010. Quality-Speed Conundrum: Trade-offs in Customer-Intensive Services. *Management Science* **57**(1) 40–56.

Andritsos, D. A., C. S. Tang. 2014. Linking process quality and resource usage: An empirical analysis. *Production and Operations Management* **23**(12) 2163–2177.

Ashraf, N., O. Bandiera, S. S. Lee. 2014. Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization* **100** 44–63.

Autor, D. 2003. Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics* **21**(1) 1–42.

Bandiera, O., I. Barankay, I. Rasul. 2013. Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association* **11** 1079–1114.

Barankay, I. 2012. Rank Incentives: Evidence from a Randomized Workplace Experiment.

Batt, R. J., C. Terwiesch. 2015. How a Service Process Adapts to Load: An Econometric Analysis of Patient Treatment in the Emergency Department.

Bendoly, E. 2013. Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management* **31**(1-2) 62–71.

Bertrand, M., E. Duflo, S. Mullainathan. 2004. How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics* **119**(1) 249–275.

Berwick, D. M. 2003. Errors today and errors tomorrow. *New England Journal of Medicine* **348**(25) 2570–2.

Beshears, J., F. Gino. 2015. Leaders as Decision Architects: Structure Your Organization's Work to Encourage Wise Choices. *Harvard Business Review* **93**(5) 52–62.

Blanes i Vidal, J., M. Nossol. 2011. Tournaments Without Prizes: Evidence from Personnel Records. *Management Science* **57**(10) 1721–1736.

Bohn, R. E. 1995. Noise and Learning in Semiconductor Manufacturing. *Management Science* **41**(1) 31–42.

Boudreau, K. J., K. R. Lakhani, M. Menietti. 2015. Performance Responses to Competition across Skill-levels in Rank Order Tournaments: Field Evidence and Implications for Tournament Design. *RAND Journal of Economics* .

Brown, M. Z., M. M. Linehan, K. A. Comtois, A. Murray, A. L. Chapman. 2009. Shame as a prospective predictor of self-inflicted injury in borderline personality disorder: A multi-modal analysis. *Behaviour Research and Therapy* **47**(10) 815–822.

Centers for Disease Control and Prevention. 2002. Guideline for Hand Hygiene in Health-Care Settings. *Morbidity and Mortality Weekly Report* **51**.

Centers for Medicare & Medicaid Services. 2002. Medicare Program; Revisions to Payment Policies Under the Physician Fee Schedule for Calendar Year 2003 and Inclusion of Registered Nurses in the Personnel Provision of the Critical Access Hospital Emergency Services Requirement for Frontier Areas and Rem. *Federal Register* **67**(251) 79972.

Chan, T. Y., J. Li, L. Pierce. 2014. Compensation and Peer Effects in Competing Sales Teams. *Management Science* **60**(8) 1965–1984.

Chandrasekaran, A., C. Senot, K. K. Boyer. 2012. Process Management Impact on Clinical and Experiential Quality: Managing Tensions Between Safe and Patient-Centered Healthcare. *Manufacturing & Service Operations Management* **14**(4) 548–566.

Charness, G., D. Masclet, M. C. Villeval. 2014. The Dark Side of Competition for Status. *Management Science* **60**(1) 38–55.

Cowgill, B. 2015. Competition and Productivity in Employee Promotion Contests.

DeHoratius, N., A. Raman. 2008. Inventory Record Inaccuracy: An Empirical Analysis. *Management Science* **54**(4) 627–641.

Dobson, G., T. Tezcan, V. Tilson. 2013. Optimal Workflow Decisions for Investigators in Systems with Interruptions. *Management Science* **59**(5) 1125–1141.

Dunn, J., N. E. Ruedy, M. E. Schweitzer. 2012. It hurts both ways: How social comparisons harm affective and cognitive trust. *Organizational Behavior and Human Decision Processes* **117**(1) 2–14.

Edelman, B., I. Larkin. 2015. Social Comparisons and Deception Across Workplace Hierarchies: Field and Experimental Evidence. *Organization Science* **26**(1) 78–98.

Frei, F. X., R. Kalakota, A. J. Leone, L. M. Marx. 1999. Process Variation as a Determinant of Bank Performance: Evidence from the Retail Banking Study. *Management Science* **45**(9) 1210–1220.

Fung, C. H., Y.-W. Lim, S. Mattke, C. Damberg, P. G. Shekelle. 2008. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Annals of Internal Medicine* **148**(2) 111–23.

Georgakopoulos, D., M. Hornick, A. Sheth. 1995. An overview of workflow management: From process modeling to workflow automation infrastructure. *Distributed and Parallel Databases* **3**(2) 119–153.

Gilboy, N., T. Tanabe, D. Travers, A. Rosenau. 2011. *Emergency Severity Index (ESI): A Triage Tool for Emergency Department Care, Version 4, Implementation Handbook*. Agency for Healthcare Research and Quality, Rockville, MD.

Gorman, A. 2015. How One Hospital Reduced Unnecessary C-Sections. URL `http://www.theatlantic.com/health/archive/2015/05/how-one-hospital-reduced-unnecessary-c-sections/392924/`.

Greene, W. 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal* **7** 98–119.

Guttmann, A., H. J. Schultz, M. J. Vermeulen, T. A. Stukel. 2011. Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada. *BMJ* **342** d2983.

Hair, J. F., R. E. Anderson, R. L. Tatham, W. C. Black. 1998. *Multivariate data analysis*. 5th ed. Prentice Hall, Upper Saddle River, NJ.

Hannan, R. L., G. P. McPhee, A. H. Newman, I. D. Tafkov. 2013. The effect of relative performance information on performance and effort allocation in a multi-task environment. *Accounting Review* **88**(2) 553–575.

Harder, D. W., S. J. Lewis. 1987. The assessment of shame and guilt. J. N. Butcher, C. D. Spielberger, eds., *Advances in personality assessment, Vol. 6*. Erlbaum, Hillsdale, NJ, 89–114.

Hopp, W. J., S. M. R. Iravani, F. Liu. 2009. Managing White-Collar Work: An Operations-Oriented Survey. *Production and Operations Management* **18**(1) 1–32.

Hopp, W. J., S. M. R. Iravani, G. Y. Yuen. 2007. Operations Systems with Discretionary Task Completion. *Management Science* **53**(1) 61–77.

Hopp, W. J., M. L. Spearman. 2000. *Factory Physics: Foundations of Manufacturing Management*. 2nd ed. McGraw-Hill, Burr Ridge, IL.

Imbens, G. W., J. M. Wooldridge. 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature* **47**(1) 5–86.

Jaikumar, R. 2005. From Filing and Fitting to Flexible Manufacturing: A Study in the Evolution of Process Control. *Foundations and Trends in Technology, Information and Operations Management* **1**(1).

KC, D. S. 2014. Does Multitasking Improve Performance? Evidence from the Emergency Department. *Manufacturing & Service Operations Management* **16**(2) 168–183.

KC, D. S., C. Terwiesch. 2009. Impact of Workload on Service Time and Patient Safety: An Econometric Analysis of Hospital Operations. *Management Science* **55**(9) 1486–1498.

Kingman, J. F. C. 1961. The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society* **57**(04) 902–904.

Klassen, R. D., L. J. Menor. 2007. The process management triangle: An empirical investigation of process trade-offs. *Journal of Operations Management* **25**(5) 1015–1034.

Kuhnen, C. M., A. Tymula. 2012. Feedback, Self-Esteem, and Performance in Organizations. *Management Science* **58**(1) 94–113.

Kuziemko, I., R. W. Buell, T. Reich, M. I. Norton. 2014. Last-place aversion: Evidence and redistributive implications. *Quarterly Journal of Economics* **129** 105–149.

Mas, A., E. Moretti. 2009. Peers at Work. *American Economic Review* **99**(1) 112–145.

McCarthy, M. L., R. Ding, J. M. Pines, C. Terwiesch, M. Sattarian, J. a. Hilton, J. Lee, S. L. Zeger. 2012. Provider variation in fast track treatment time. *Medical Care* **50**(1) 43–9.

McClelland, M. S., D. Lazar, V. Sears, M. Wilson, B. Siegel, J. M. Pines. 2011. The Past, Present, and Future of Urgent Emergency Department Flow Improvement. *Academic Emergency Medicine* **18** 1392–1399.

McGregor, J. 2006. The Struggle To Measure Performance. *Bloomgberg Businessweek* .

Moran, S., M. E. Schweitzer. 2008. When Better Is Worse: Envy and the Use of Deception. *Negotiation and Conflict Management Research* **1**(1) 3–29.

Morganti, K. G., S. Bauhoff, J. C. Blanchard, M. Abir, N. Iyer, A. Smith, J. V. Vesely, E. N. Okeke, A. L. Kellermann. 2013. The Evolving Role of Emergency Departments in the United States. Tech. rep., RAND Corporation, Santa Monica, CA. URL `http://www.rand.org/pubs/research_reports/RR280`.

Nagin, D. S., J. B. Rebitzer, S. Sanders, L. J. Taylor. 2002. Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review* **92**(4) 850–873.

Narayanan, S., S. Balasubramanian, J. M. Swaminathan. 2009. A Matter of Balance: Specialization, Task Variety, and Individual Learning in a Software Maintenance Environment. *Management Science* **55**(11) 1861–1876.

Netessine, S., V. Yakubovich. 2012. The Darwinian Workplace. *Harvard Business Review* **90**(5) 25–28.

Nordstrom, R., P. Lorenzi, R. V. Hall. 1991. A Review of Public Posting of Performance Feedback in Work Settings. *Journal of Organizational Behavior Management* **11**(2) 101–124.

Ohno, T. 1988. *Toyota Production System: Beyond Large-Scale Production*. Productivity Press, Portland, OR.

Patel, P. B., D. R. Vinson. 2005. Team Assignment System: Expediting Emergency Department Care. *Annals of Emergency Medicine* **46** 499–506.

Pierce, L., D. Snow, A. McAfee. 2015. Cleaning House : The Impact of Information Technology Monitoring on Employee Theft and Productivity. *Management Science* .

Pitts, S. R., E. R. Carrier, E. C. Rich, A. L. Kellermann. 2010. Where Americans get acute care: Increasingly, it's not at their doctor's office. *Health Affairs* **29**(9) 1620–1629.

Preacher, K. J., A. F. Hayes. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* **40**(3) 879–891.

Pronovost, P. J., D. Needham, S. M. Berenholtz, D. Sinopoli, H. Chu, S. Cosgrove, B. Sexton, R. Hyzy, R. Welsh, G. Roth, J. Bander, J. Kepros, C. Goeschel. 2006. An intervention to decrease catheter-related bloodstream infections in the ICU. *New England Journal of Medicine* **355**(26) 2725–32.

Roels, G., X. Su. 2014. Optimal Design of Social Comparison Effects: Setting Reference Groups and Reference Points. *Management Science* **60**(3) 606–627.

Schull, M., M. Vermeulen, A. Guttmann, T. Stukel. 2015. Better performance on length-of-stay benchmarks associated with reduced risk following emergency department discharge: an observational cohort study. *Canadian Journal of Emergency Medicine* **17**(3) 253–262.

Schultz, K. L., D. C. Juran, J. Boudreau. 1999. The Effects of Low Inventory on the Development of Productivity Norms. *Management Science* **45**(12) 1664–1678.

Shadish, W. R., T. D. Cook, D. T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shah, R., S. M. Goldstein, B. T. Unger, T. D. Henry. 2008. Explaining Anomalous High Performance in a Health Care Supply Chain. *Decision Sciences* **39**(4) 759–789.

Soremekun, O. A., C. Terwiesch, J. M. Pines. 2011. Emergency medicine: An operations management view. *Academic Emergency Medicine* **18**(12) 1262–1268.

Spaite, D. W., F. Bartholomeaux, J. Guisto, E. Lindberg, B. Hull, A. Eyherabide, S. Lanyon, E. A. Criss, T. D. Valenzuela, C. Conroy. 2002. Rapid process redesign in a university-based emergency department: Decreasing waiting time intervals and improving patient satisfaction. *Annals of Emergency Medicine* **39**(2) 168–177.

Spear, S., H. K. Bowen. 1999. Decoding the DNA of the Toyota Production System. *Harvard Business Review* **77**(5) 96–106.

Staats, B. R., D. J. Brunner, D. M. Upton. 2011. Lean principles, learning, and knowledge work: Evidence from a software services provider. *Journal of Operations Management* **29**(5) 376–390.

Staats, B. R., H. Dai, D. Hofmann, K. L. Milkman. 2015. Motivating Process Compliance through Individual Electronic Monitoring: An Empirical Examination of Hand Hygiene in Healthcare.

Staats, B. R., F. Gino. 2012. Specialization and Variety in Repetitive Tasks: Evidence from a Japanese Bank. *Management Science* **58**(6) 1141–1159.

Szulanski, G. 1996. Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm. *Strategic Management Journal* **17**(Winter Special Issue) 27–43.

Tafkov, I. D. 2013. Private and public relative performance information under different compensation contracts. *Accounting Review* **88**(1) 327–350.

Tan, T. F., S. Netessine. 2014. When Does the Devil Make Work? An Empirical Study of the Impact of Workload on Worker Productivity. *Management Science* **60**(6) 1574–1593.

Taylor, F. W. 1911. *The Principles of Scientific Management*. Harper & Brothers, New York, NY.

Tucker, A. L., I. M. Nembhard, A. C. Edmondson. 2007. Implementing New Practices: An Empirical Study of Organizational Learning in Hospital Intensive Care Units. *Management Science* **53**(6) 894–907.

Tucker, A. L., S. J. Spear. 2006. Operational failures and interruptions in hospital nursing. *Health Services Research* **41**(3 I) 643–662.

Vanek Smith, S. 2015. I, Waiter. URL `http://www.npr.org/templates/transcript/transcript.php?storyId=407086723`.

Wagner, A. K., S. B. Soumerai, F. Zhang, D. Ross-Degnan. 2002. Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics* **27**(4) 299–309.

Wang, L., I. Gurvich, J. A. V. Mieghem, K. J. O'Leary. 2015. Collaboration and Professional Labor Productivity : An Empirical Study of Physician Workflows in a Hospital.

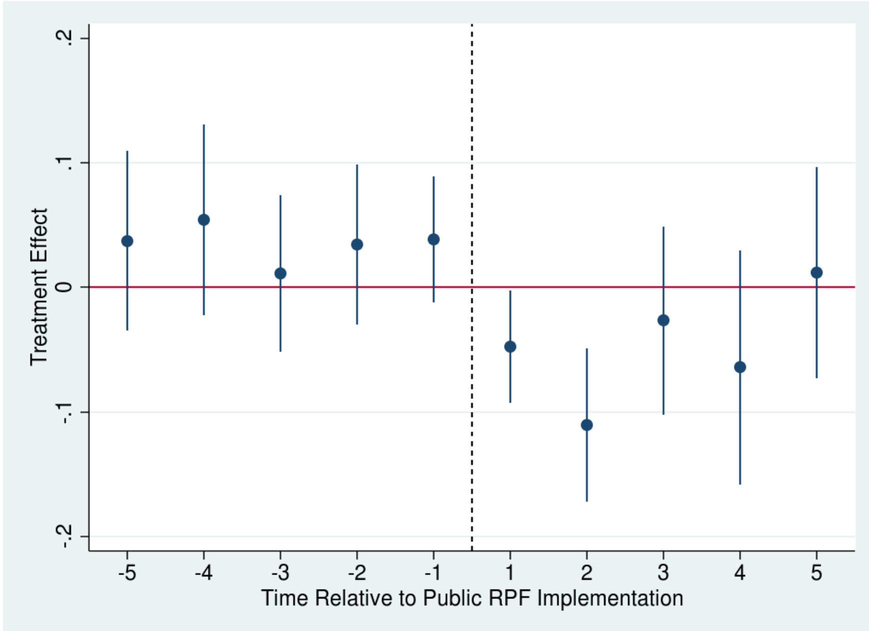
Watson, D., L. A. Clark, A. Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* **54**(6) 1063–1070.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data: Second Edition*. 2nd ed. MIT Press, Cambridge, Massachusetts.

Wooldridge, J. M. 2012. *Introductory Econometrics: A Modern Approach*. 5th ed. South-Western Cengage Learning, Mason, OH.

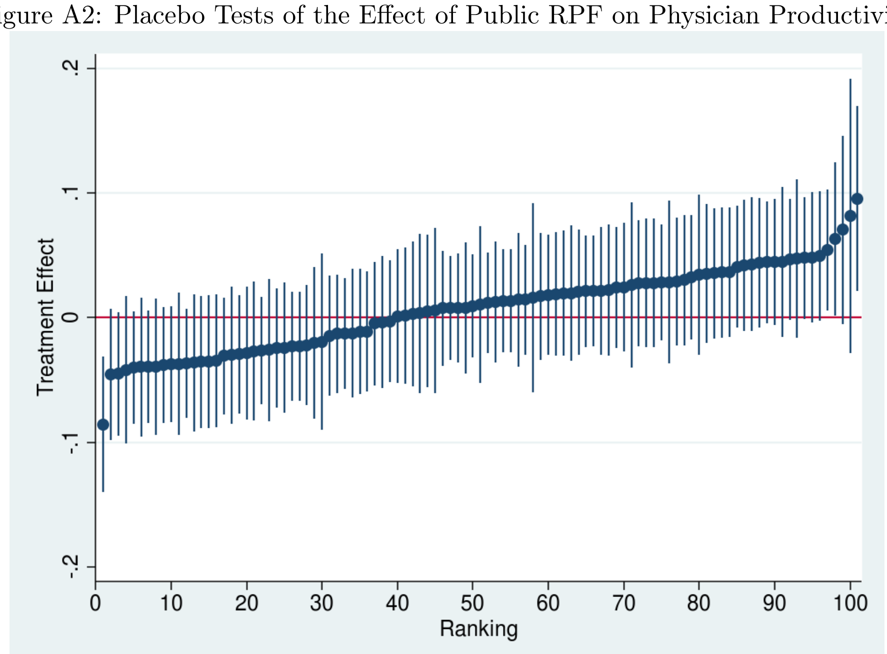
Zizzo, D. J. 2002. Between utility and cognition: The neurobiology of relative position. *Journal of Economic Behavior and Organization* **48**(1) 71–91.

Figure A1: Leads and Lags Model Testing Pre-intervention Time Trends



Notes. Each point and its whiskers (95% confidence intervals) correspond to an estimation of equation (1) as described in section 4.1 that replaces the intervention indicator *Post* with five lead variables and five lag variables. Lead and lag variables correspond to each of the 5 quarters before and after the intervention, respectively. The vertical dashed line indicates the time when public RPF was implemented at Treatment ED (August 2010).

Figure A2: Placebo Tests of the Effect of Public RPF on Physician Productivity



Notes. Placebo intervention dates were randomly assigned 100 times. Each point and its whiskers (95% confidence intervals) correspond to an estimation of equation (1) as described in section 4.1. The coefficient and confidence interval for the true data are represented on the far left (rank 1).

Table A1: System-level Effects of Public RPF on ED Length of Stay

Variables	(1) Logged ED LOS Excluding Boarding Time	(2) Logged ED LOS Including Boarding Time	(3) Logged Throughput Rate
Post X Treat	-0.100***	-0.113***	0.026*
	(0.025)	(0.025)	(0.013)
ESI level 2	0.343***	0.296***	0.000
	(0.021)	(0.022)	(0.004)
ESI level 3	0.021	-0.054*	0.034***
	(0.024)	(0.023)	(0.005)
Age	0.007***	0.008***	-0.001***
	(0.000)	(0.000)	(0.000)
Female	0.032***	0.024***	0.001
	(0.003)	(0.003)	(0.001)
Total patients	0.023***	0.023***	0.013***
	(0.002)	(0.002)	(0.001)
$(Total\ patients)^2$	-0.001***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)
MDs on shift	-0.003	-0.001	-0.008***
	(0.004)	(0.004)	(0.001)
Afternoon shift	-0.029*	-0.030*	0.009*
	(0.012)	(0.012)	(0.004)
Overnight shift	0.025	0.034	0.004
	(0.020)	(0.020)	(0.005)
Time-varying controls	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
Physician FE	No	No	Yes
Observations	274,339	278,676	196,326
Adjusted R-squared	0.122	0.147	0.211

Notes. Regressions are fixed effects log-linear difference-in-differences models estimated at the patient encounter level. Throughput rate is operationalized as the number of patients discharged by a physician on a given shift, adjusted for the duration of the shift. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p <0.05, **p <0.01, ***p <0.001.

Table A2: Alternate Model Specifications

Variables	Restricted Time Frame (1) 3 Months Pre/Post	(2) 6 Months Pre/Post	(3) 12 Months Pre/Post	Other (4) No Washout Period	(5) Linear DV	(6) Interrupted Time Series
Post X Treat	-0.067**	-0.105***	-0.095***	-0.083**	-16.657***	
	(0.026)	(0.026)	(0.028)	(0.027)	(2.807)	
ESI level 2	0.352***	0.324***	0.273***	0.275***	72.915***	0.440***
	(0.058)	(0.042)	(0.033)	(0.030)	(6.219)	(0.034)
ESI level 3	-0.021	-0.061	-0.129***	-0.117***	1.648	0.001
	(0.057)	(0.043)	(0.036)	(0.032)	(5.642)	(0.047)
Age	0.011***	0.011***	0.011***	0.011***	1.483***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.022)	(0.000)
Female	0.055***	0.049***	0.048***	0.048***	0.345	0.030***
	(0.008)	(0.006)	(0.005)	(0.005)	(1.001)	(0.007)
Total patients	0.010*	0.007*	0.005*	0.005*	1.758***	0.001
	(0.005)	(0.003)	(0.002)	(0.002)	(0.471)	(0.002)
$(Total\ patients)^2$	-0.001*	-0.000*	-0.000*	-0.000**	-0.088***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.025)	(0.000)
MDs on shift	-0.004	-0.006	-0.004	-0.004	-0.499	-0.010***
	(0.004)	(0.003)	(0.003)	(0.002)	(0.380)	(0.003)
Afternoon shift	-0.038**	-0.013	-0.022*	-0.031***	-1.112	-0.023*
	(0.013)	(0.010)	(0.009)	(0.007)	(1.169)	(0.012)
Overnight shift	-0.001	0.015	-0.002	-0.014	5.429**	-0.004
	(0.019)	(0.014)	(0.012)	(0.010)	(1.854)	(0.019)
Post						-0.088**
						(0.027)
Time						-0.002***
						(0.000)
Time after intervention						0.001
						(0.001)
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45,222	91,123	181,935	269,270	261,586	126,196
Adjusted R-squared	0.174	0.179	0.189	0.189	0.097	0.198

Notes. Columns 1 - 4 are fixed effects log-linear difference-in-differences models, Column 5 is a fixed effects linear difference-in-differences model, and Column 6 is a fixed effects log-linear interrupted time series model. All models are estimated at the patient encounter level. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. **p* <0.05, ***p* <0.01, ****p* <0.001.

Table A3: Heterogeneous Effect of Public RPF on Physician Productivity with Alternate Thresholds

Variables	(1) Top/Bottom 2	(2) Top/Bottom 3	(3) Top/Bottom 5
Post X Treat	-0.082**	-0.067*	-0.072*
	(0.029)	(0.027)	(0.030)
Post X Treat X Top	0.002	-0.237	-0.176
	(0.031)	(0.178)	(0.094)
Post X Treat X Bottom	-0.084**	-0.049	-0.022
	(0.029)	(0.035)	(0.034)
ESI level 2	0.270***	0.270***	0.270***
	(0.033)	(0.033)	(0.033)
ESI level 3	-0.124***	-0.123**	-0.123**
	(0.037)	(0.037)	(0.037)
Age	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)
Female	0.047***	0.047***	0.047***
	(0.005)	(0.005)	(0.005)
Total patients	0.005	0.005	0.005
	(0.002)	(0.002)	(0.002)
$(Total\ patients)^2$	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)
MDs on shift	-0.005	-0.005	-0.005
	(0.002)	(0.002)	(0.002)
Afternoon shift	-0.035***	-0.035***	-0.034***
	(0.008)	(0.008)	(0.008)
Overnight shift	-0.018	-0.019	-0.018
	(0.011)	(0.010)	(0.010)
Time-varying controls	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes
Observations	232,297	232,297	232,297
Adjusted R-squared	0.181	0.181	0.181

Notes. Regressions are fixed effects log-linear difference-in-differences models estimated at the patient encounter level. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p <0.05, **p <0.01, ***p <0.001.

Table A4: Average Effect of Public RPF on Physician Productivity with Various Exclusions

Variables	(1) Excluding Patients With LOS > 48 hours	(2) Excluding Patients Receiving Psychiatry Consult	(3) Excluding ESI Level 1 Patients	(4) Excluding Expired Patients	(5) Excluding LWBS, AMA, Eloped Patients
Post X Treat	-0.085**	-0.084**	-0.086**	-0.086**	-0.086**
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
ESI level 2	0.268***	0.199***	0.393***	0.268***	0.270***
	(0.029)	(0.028)	(0.014)	(0.029)	(0.029)
ESI level 3	-0.120***	-0.114***		-0.125***	-0.124***
	(0.033)	(0.032)		(0.033)	(0.032)
Age	0.011***	0.011***	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.048***	0.052***	0.047***	0.048***	0.047***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Total patients	0.006*	0.005*	0.006*	0.006*	0.006*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$(Total\ patients)^2$	-0.000**	-0.000**	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MDs on shift	-0.005*	-0.004	-0.005*	-0.005*	-0.005*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Afternoon shift	-0.032***	-0.037***	-0.032***	-0.032***	-0.032***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Overnight shift	-0.015	-0.027**	-0.015	-0.015	-0.015
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Time-varying controls	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes
Observations	261,481	255,942	260,238	261,569	260,944
Adjusted R-squared	0.191	0.195	0.191	0.190	0.191

Notes. LWBS = Left without being seen. AMA = Against medical advice. Regressions are fixed effects log-linear difference-in-differences models estimated at the patient encounter level. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p <0.05, **p <0.01, ***p <0.001.