The Impact of Batching Advanced Imaging Tests in Emergency Departments

Jacob Jameson
Harvard Kennedy School, Harvard University,

Soroush Saghafian
Harvard Kennedy School, Harvard University,

Robert Huckman
Harvard Business School, Harvard University,

Nicole Hodgson
Department of Emergency Medicine, Mayo Clinic of Arizona,

Joshua Baugh
Department of Emergency Medicine, Massachusetts General Hospital,

Using detailed electronic health record data from two major U.S. emergency departments (EDs), we use practice variation across physicians to uncover the operational impact of batch ordering imaging tests. Using causal inference approaches, we find that assignment of an arriving patient to an ED physician who is a "batcher" versus a "sequencer" causally increases patient length of stay, time to disposition, and the number of imaging tests ordered. We find evidence that the impact of batching on length of stay is heavily mediated by the additional imaging tests as well as probability of being admitted to the hospital post ED service, suggesting this ordering strategy may lead to more conservative clinical decision-making and critical bottlenecks in patient flow. Conversely, sequencing imaging tests by ED physicians poses an "information gain" advantage compared to batching: the information obtained from a prior test allows eliminating the need for ordering some future tests. Put together, our findings indicate that batch ordering may not be an optimal strategy for managing diagnostic imaging in emergency care, and that interventions to reduce physician discretion and behavioral tendencies in batch ordering may be warranted.

Key words: Emergency Department operations; Diagnostic imaging; Batch ordering; Physician practice patterns; Patient outcomes; Health care efficiency

1. Introduction

Advanced imaging is simultaneously an emergency department's (ED) most powerful diagnostic tool and one of their greatest operational bottlenecks (Rogg et al. 2017). The use of advanced diagnostic imaging has risen dramatically over the past two decades, transforming from a limited resource to a cornerstone of emergency care (Juliusson et al. 2019, Smith-Bindman et al. (2019)). Yet this transformation has intensified operational challenges in EDs, where the complexity stems from multiple modifiable and non-modifiable constraints: limited equipment availability, complex scheduling requirements across different modalities, extended wait times for both image acquisition and interpretation, and growing backlogs in radiologist reading queues. Because of this, diagnostic imaging represents one of the most resource-intensive and operationally complex components of ED care (Mills et al. 2015, Baloescu (2018), Poyiadji et al. (2023)).

Diagnostic imaging is not without consequences for patients. Imaging bottlenecks can significantly impact ED length of stay (LOS) (Cournane et al. 2016), and undergoing advanced imaging may expose patients to higher costs, increased exposure to radiation, increased incidental findings that may lead to unnecessary follow-up testing, contrast-induced nephropathy, and contrast-induced allergic reactions (Valtchinov et al. 2019, Raja et al. (2014)). Given these risks and resource constraints, efficient management of diagnostic testing is critical for both patient outcomes and ED operations (Balogh et al. 2015). Physicians have considerable discretion over how they order these tests, despite wide variation in diagnostic test ordering behavior being well documented (Miller 1994, Solomon et al. (1998), Wennberg (1984), Daniels and Schroeder (1977)). Yet, little is known about how test ordering strategies should be managed when this discretion exists.

We consider an ED physician's decision to order imaging tests for their patient as an optimization problem, where the physician must balance the trade-offs between the advantages of ordering multiple tests simultaneously (batch) at the start of the patient encounter (e.g., expediting the diagnostic process if the tests are eventually necessary for diagnosis and disposition) and its disadvantages (e.g., increasing the total time spent in the ED because of unnecessary additional tests) (Tamburrano et al. 2020, Perotte et al. (2018), Lyu et al. (2017), Traub et al. (2018)).

Batch ordering stands in contrast to the more common practice of ordering tests sequentially, where physicians order one test at a time, review the results, and then decide whether to order additional tests based on the information obtained from the previous test. While, in theory, sequential ordering may serve as a natural filter to prevent unnecessary testing, it can also result in longer delays compared to batching when multiple tests are ultimately needed. This tension between potential efficiency gains and the risk of unnecessary testing is particularly important given growing concerns about imaging overutilization in emergency care (Baloescu 2018, Mills et al. (2015)). Beyond avoiding delays, the decision to batch should also be made by considering both immediate operational implications and broader quality of care considerations (Feizi et al. 2023). This decision is inherently complex, requiring physicians to weigh time-sensitive diagnostic needs against resource constraints and the risk of ordering tests that may prove unnecessary once earlier results become available.

In this paper, we explore the causal effects of batch ordering advanced imaging tests— a practice in which physicians order multiple tests simultaneously— on operational performance and patient outcomes in the ED. Specifically, we focus on batch orders that include different types of imaging tests because the operational implications of batching these tests differ as they cannot be run in a single scanning session. In collaboration with two leading US hospitals, we start by providing evidence on a significant level of variations in the batching behavior of emergency physicians. We find that across physicians who work in the same ED, treating patients of the same complaint and severity, batching varies significantly (Figure 1). By making use of this variation, we next investigate whether being seen by a "batcher" or a "sequencer" physician has important implications in terms of performance metrics such as length of stay (LOS), number of imaging studies, and disposition. Finally, we shed light on circumstances under which physicians are more likely to batch order imaging tests.

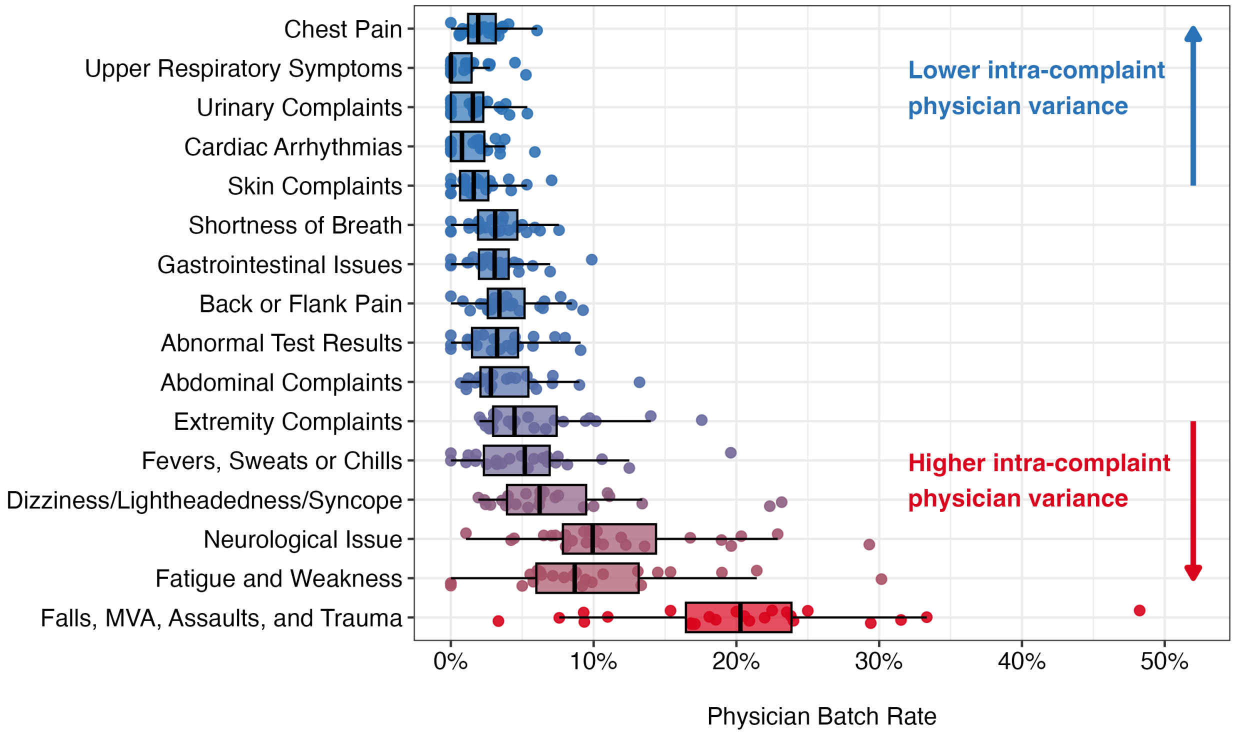

Figure 1 Physician Variation in Batch Ordering Imaging Tests

Notes: This figure highlights the marked differences among Mayo Clinic ED physicians in their propensity to batch order imaging tests. Batch rates are crude rates calculated by dividing the number of patient encounters where the physician batch ordered imaging tests for a complaint by the number of patient encounters they had with that complaint.

1.1. Challenges and Empricial Strategy

While the importance of understanding batching behavior in EDs is clear, generating causal evidence on its effects presents several empirical challenges. First, physicians' decisions to batch order tests are inherently endogenous. For example, the choice to batch may be correlated with unobservable patient characteristics, physician workload, or ED conditions that independently affect outcomes. Second, studying the timing and sequencing of diagnostic testing requires granular operational data that captures precise timestamps of test orders, completions, and clinical decisions, many of which is not information not available in traditional claims databases. Third, while prior studies have leveraged quasi-random patient assignment in EDs (Eichmeyer and Zhang 2022, Gowrisankaran et al. (2022), Coussens and Ly (2024)), establishing true quasi-randomization requires detailed understanding of institutional assignment mechanisms that may vary across settings. Thus, providing causal claims that can be generalizable beyond a single ED is not straightforward.

Our study addresses these challenges through several unique features. First, we utilize detailed electronic health record data from two leading U.S. hospitals (Mayo Clinic and Massachusetts General Hospital (MGH)) that capture the complete temporal sequence of clinical decisions, including exact timestamps of test orders, results availability, and disposition decisions (Table 1). This granularity allows us to precisely measure both batch ordering behavior and its downstream effects on patient flow. Second, our primary analysis leverages Mayo Clinic's rotational patient assignment system, where we can verify that patients are randomly assigned to physicians via a round-robin algorithm that does not consider patient characteristics or physician workload (Traub et al. 2016). We make use of this random assignment at the Mayo clinic to gain deeper insights into the causal impacts of being seen by a "batcher." We then validate our findings using data from MGH, which employs a completely different and non-random patient-to-physician assignment. We do so by creating suitable controls, which enables constructing a quasi-random assignment mechanism, enabling us to investigate the validity of our main findings in two vastly different study sites.

Our empirical strategy closely follows the literature that relies on quasi-random assignment of agents to cases, often referred to as the "judges design." (Dahl et al. 2014a, Dobbie et al. (2018)). Studies in this literature typically exploit variation in the sentencing leniency of judges who work in the same court. Similarly, we exploit batching variation across physicians who work in the same ED through a measure we define as "batch tendency." In its reduced form, under the assumption of random or quasi-random assignment, this approach allows us to identify the causal effect of being assigned to different types of physicians (i.e., batcher or sequencer). Under additional assumptions, an instrumental variable (as we will define it) allows us to estimate the causal effect of the decision to batch on various important ED measures.

Our instrumental variables approach provides us with the Local Average Treatment Effect (LATE) of batch ordering, which is the effect of batch ordering on the subset of patients whose batching decision changes due to the instrument. This is particularly important in the context of

Table 1 Descriptive Statistics of Emergency Department Encounters

Variable	Mayo Clinic (Median [IQR])	MGH (Median [IQR])
Patients encounters[a]	n = 48,854	n = 111,710
Panel A. Patient Severity		
Tachycardic	19.2%	21.3%
Tachypneic	8.8%	5.4%
Febrile	2.2%	1.5%
Hypotensive	1.4%	1.0%
Emergency Severity Index	2.8 [2, 3]	2.8 [2, 3]
Panel B. Patient Demographics		
Male	46.5%	51.3%
Race: White	88.4%	61.3%
Race: Black	4.2%	12.3%
Race: Asian	3.05%	4.89%
Arrival age	57.7 [43, 74]	49.6 [32, 65]
Panel C. Diagnostic Tests and Outcomes		
X-ray performed	43.3%	40.1%
Ultrasound performed	11.3%	18.3%
Non-contrast CT performed	35.5%	30.5%
Contrast CT performed	17.7%	13.0%
MRI performed[b]	—	6.3%
Labs ordered	73.7%	84.9%
Time from arrival to triage (mins)	8.0 [4, 10]	12.2 [3, 13]
LOS (min)	246 [152, 306]	426 [241, 1006]
Patient discharged	66.8%	65.1%
Patient admitted	18.7%	22.8%
Patients revisited within 72 hours	3.8%	3.1%
Order to result[c]: X-Ray (mins)	67.2 [36, 79]	63.4 [31.4, 111.7]
Order to result: Ultrasound (mins)	165 [71, 150]	119 [71, 213]
Order to result: Contrast CT (mins)	142 [86, 153]	167.3 [112.0, 260.6]
Order to result: Non-Contrast CT (mins)	89.7 [50, 102]	185.3 [116.0, 290.2]
Order to result: MRI (mins)	—	374.7 [229.5, 683.7]

Notes: This table reports summary statistics for emergency department visits during the study period. Values are presented as mean (IQR) when available. Vital signs were categorized as follows: tachycardia (pulse more significant than 100), tachypnea (respiratory rate greater than 20), fever (temperature greater than 38°C), and hypotension (systolic blood pressure less than 90).

[a] The number of patient encounters is calculated as the number of unique patient visits during the study period.

[b] MRI data is only available for MGH.

[c] Order to result times are calculated as the difference between the time the test was ordered and the time the result was available in the electronic health record. This does not account for the time it takes for the radiologist to review the results.

the ED, where there are clear cases where batch ordering is necessary (e.g., when multiple tests are essential and almost every physician would batch) and cases where it is not (e.g., when one or no tests are needed and almost no physician would batch). By estimating the effect for the marginal patients whose tests may or may not be batched depending on the provider, we generate deeper insights into the important implications of batch ordering as a result of physician discretion.

1.2. Main Findings and Contributions

Despite the conventional wisdom that batching tests in the ED enables physicians to gain efficiency, we find that batch ordering imaging tests negatively impacts operational metrics. In particular, our results show that the marginal batched patient experiences an 85.8% increase in total ED LOS and an 85.4% increase in time to disposition compared to patients who have their tests ordered sequentially. Furthermore, when physicians batch order imaging tests, they order 1.4 additional test per patient encounter compared to the sequential strategy. These effects remain robust even after adjusting for patient characteristics, ED conditions, and physician experience. Through mediation analysis, we find that the increase in testing volume is a key driver of the increased LOS associated with batch ordering, and that this effect of batching on LOS is further mediated by admission.

Examining the drivers of batch ordering, we find that physicians are more likely to batch order tests for patients with higher acuity levels and more complex chief complaints. This batching tendency, however, varies systematically across physicians and persists even after controlling for patient mix and ED conditions. Finally, we find that ED crowding significantly reduces the negative impact of batching behavior. For example, we observe that during periods of major overcapacity, the effect of batching on testing volume drops by nearly half compared to normal operations, suggesting that resource constraints may induce batchers to be more selective and exclude unnecessary tests from their batches.

Taken together, our results indicate that despite the perceived workflow advantages of initiating multiple diagnostic processes simultaneously, batch ordering leads to significantly longer processing times and increased resource utilization without corresponding improvements in patient outcomes. These findings have important implications, highlighting that ED managers should think about strategies to reduce physician discretion in test ordering as a way of improving the diagnostic testing workflows.

2. Related Literature

In EDs, physicians face a fundamental choice in how they sequence diagnostic imaging tests: they can either order multiple tests simultaneously (“batch ordering”) or order them sequentially based on progressive information gathering. Batch ordering occurs when a physician orders a comprehensive set of diagnostic tests at the start of a patient encounter, typically covering a broad range of potential diagnoses. This contrasts with sequential ordering, where tests are ordered one at a time with each subsequent test decision informed by prior test results.

2.1. Physician Decision-Making Under Constraints:

ED physicians operate under significant time pressure and workload, which substantially influences their decision-making processes. Workload management and carefully balancing the tradeoffs between speed and quality of care are of high importance for physicians making decisions in a complex and multitasking environment such as the ED (Saghafian et al. 2018, Leppink and Hanham (2019)) and leads to multiple documented effects on physician performance: task switching (KC 2013, Skaugset et al. (2016)), increased cognitive load (KC and Terwiesch 2009, Pines (2017)), increased stress (Chisholm et al. 2000, Bendoly (2011)), and increased interruptions (Chisholm et al. 2000, Chisholm et al. (2001)).

Prior studies have identified several mechanisms that may alter physicians' test-ordering behavior under increased workload. First, physicians may order more diagnostic tests as an alternative to direct patient contact when they have less time to spend with each patient (Batt and Terwiesch 2016). Second, increased stress and interruptions can hinder systematic decision-making and critical thinking (Chisholm et al. 2000). Because test ordering requires less critical thinking compared to diagnosis through direct patient contact (Pines 2009), physicians may order more tests when workload increases or when their queue of patients is longer. Third, physicians may order additional tests to temporarily reduce workload while waiting for test results (Berry Jaeker and Tucker 2020). By ordering multiple tests upfront, physicians can defer complex diagnostic reasoning until all results are available, potentially reducing the cognitive strain of repeated task-switching and decision-making under uncertainty. A cognitive load theory of batching would suggest batching may be more prevalent during periods of high workload or complexity.

Recent work has documented significant variation in ED physician testing and admitting practices (Hodgson et al. 2018, Coussens and Ly (2024), Smulowitz et al. (2021)). This variation extends to batch ordering practices as well, where physicians differ systematically in their propensity to order multiple imaging tests simultaneously (Jameson et al. 2024). However, the drivers and operational impacts of this variation remain a significant gap in the literature. Our study aims to fill this gap by examining the factors that drive batch ordering behavior and exploiting the physician variation to identify the causal effects of batch ordering on ED operations and patient outcomes.

2.2. Discretionary Behavior and Task Scheduling in Healthcare Operations

Recent work has examined how operational factors influence physicians' discretionary behavior in test ordering, revealing that decisions about diagnostic intensity are shaped by multiple operational pressures (Soltani et al. 2022). Studies have found that test utilization varies with peer observation

(Song et al. 2017), workload (Deo and Jain 2019), and the presence of justification requirements (Berry Jaeker and Tucker 2020). While additional tests can improve diagnostic accuracy, they also extend ED LOS and potentially exacerbate congestion (Chan 2018). This tension is particularly acute in imaging decisions, where test sequencing can significantly impact both patient flow and resource utilization (Cournane et al. 2016).

The decision to batch order tests represents a specific form of discretionary task ordering that has received limited attention in healthcare operations. While prior work has examined discretionary task ordering in other contexts (Ibanez et al. 2018, Ibanez and Toffel (2020)), the unique constraints of ED imaging— including capacity limitations, varying processing times across modalities, and the inability to run different imaging types simultaneously— make these decisions particularly consequential. A growing literature examines how workers exercise discretion over task ordering to improve system performance (van Donselaar et al. 2010, Campbell and Frei (2011)), but can sometimes lead workers to "choose the wrong task operationally" (Boudreau et al. 2003).

The implications of batch ordering also connect to broader theoretical work on task scheduling in resource-constrained environments. While batching strategies are often used to reduce setup times and improve throughput in manufacturing settings (Fowler and Mönch 2022), applying these principles to healthcare operations introduces unique complexities. While batching may streamline the diagnostic process by initiating multiple diagnostic processes simultaneously (Song et al. 2017), recent evidence suggests it may lead to increased testing volumes that could overwhelm imaging departments and extend wait times (Jessome 2020, Saghafian et al. (2015)). The information value of sequential testing— where results from initial tests can inform the necessity of subsequent ones— creates a fundamental tension between operational efficiency and diagnostic efficiency that has not been well-studied. Our analysis provides novel evidence on this tradeoff, showing how different test ordering strategies affect both operational metrics and clinical decision-making. Our study advances this literature by first providing the first causal evidence on how physicians' test sequencing decisions affect ED performance. Furthermore, we identify specific mechanisms through which batch ordering affects operational performance, allowing us to distinguish between efficiency gains from parallel processing and potential losses from increased diagnostic intensity.

3. Setting, Data, and Models

3.1. Empirical Setting

Our study uses data from two large U.S. emergency departments (EDs): the Mayo Clinic of Arizona and Massachusetts General Hospital (MGH). The MGH dataset, which includes 129,489 patient encounters from November 10, 2021 through December 10, 2022, provides a robust sample for validating the generalizability of our findings. However, our primary analysis focuses on the

Mayo Clinic data due to its unique feature of random patient-physician assignment, where arriving patients are randomly assigned to physicians using a rotational-basis algorithm (Traub et al. 2016), which allows for stronger causal inference. A computer algorithm randomly assigned arriving patients to physicians in a round-robin manner, where assignments were made purely rotationally without considering patient demographics, chief complaint, ESI, physician-patient load, or acuity of patients recently assigned to the physician. In essence, physician-to-patient matching can be deemed random by controlling for patient arrival time and physician shift-level variation.

The data contain information on the timing of test orders, the timing of test results, and the timing of patient disposition, among various other important triage metrics and demographic features. We focus on imaging tests (x-rays, contrast CT, non-contrast CT, ultrasound) because unlike laboratory tests, imaging tests cannot be simultaneously run due to different equiptment and settings. Therefore, the operational implications of batch ordering imaging tests are more pronounced. We do not include MRI in our analysis due to Mayo Clinic's policy of requiring inpatient admission for MRI orders if urgent or ordered as an outpatient by an outpatient physician if not urgent, therefore resulting in an extremely low ED MRI rate. MGH does not have this policy, and we therefore run the analysis with both MRI included and not included in the generalizability analysis.

3.2. Data

Our primary data comes from the ED of the Mayo Clinic of Arizona, a tertiary care hospital without obstetrical services, an inpatient pediatrics unit, or a trauma designation. During the study period from October 6, 2018, through December 31, 2019, the ED recorded 48,854 visits per year, managed across 26 treatment rooms and up to 9 hallway spaces. The department is exclusively staffed by board-eligible or board-certified emergency physicians (EPs), a rare but ideal setting for our study. Many EDs are staffed by a mix of EPs and non-EPs, both of which are responsible for ordering tests, which may introduce confounding factors. The Mayo Clinic ED is unique in that only EPs are allowed to order tests, which eliminates the potential for confounding by provider type.

We conducted a retrospective review of the comprehensive ED operational data, coinciding with the initiation of a new electronic medical record. The data includes detailed patient demographics, chief complaints, vital signs, emergency severity index (ESI), LOS, timestamps, and resource utilization metrics. This period was chosen to provide a robust data set while excluding the influence of the coronavirus pandemic. The data is summarized in Table 1. Hourly patient arrival rates to the ED are shown in Appendix Figure A1. LOS is measured from time of arrival to departure from the ED. During times of boarding, the Mayo Clinic ED will convert ED beds into temporary inpatient beds, making the endpoint for LOS for ED boarders when they are moved from their ED bed to their assigned "inpatient" bed within the ED [1].

Our research design focuses on adults who visit the Mayo Clinic of Arizona ED. We observe $48,854$ such visits during the study period. To improve power, we drop encounters with rare "reasons for visit" (RVF) (< 1000 total encounters of this kind) and complaints where a batch order occurs less than 5% of the time or no imaging is ordered. Since batch orders are rare for these cases, our physician batch tendency instrument could suffer from a weak instrument problem if we included them. Examples of complaints dropped include Skin Complaints and Urinary Complaints, as well as other complaints where multiple modalities of imaging is unlikely to occur. Excluding these conditions does not introduce selection bias unless physician test batching tendency is orthogonal to physician diagnosing behavior. While this assumption may be violated if we were to use a very detailed level of chief complaint information upon which to base our exclusion criterion, it is plausibly satisfied when using broad complaint categories as we do. In order to estimate a precise measure of physician-level batch tendency, we further restrict our sample to the $11,404$ encounters involving full-time physicians who treat over 500 ED visits per year. Our final sample includes complaints from the following categories: Neurological Issue, Abdominal Complaints, Fevers, Sweats or Chills, Falls, Motor Vehicle Crashes, Assaults, and Trauma, Dizziness/Lightheadedness/Syncope, Extremity Complaints, and Fatigue and Weakness.

3.2.1. Treatment Variable Our treatment variable in the IV analysis, $Batched_{i,t}$, is an indicator taking the value of 1 if patient i has their tests batch ordered during their ED encounter which took place on date t, and 0 otherwise. Batching occurs when a physician simultaneously orders a comprehensive set of diagnostic tests, typically covering a broad range of potential diagnoses. This contrasts with sequential ordering, where tests are ordered in sequence based on the information obtained from subsequent tests as needed.

We define "batching" in line with standard emergency medicine practices, and focus on batches that include two or more different imaging tests ordered within a 5-minute window at the start of a patient encounter. Sensitivity analyses around this time window (5-minute), batch size threshold (two or more), and when the batch occurs (beginning the encounter) show that our results are robust to variation in these values Each imaging modality, such as X-ray, contrast CT scan, non-contrast CT, and ultrasound, is considered a separate and distinct test for our study. In particular, we focus on batching instances in which the physician orders different imaging tests, because such tests cannot be done in a single scanning session (due to differences in equipment and setting).

---

[1] The Mayo Clinic hospital operations team views ED crowding and boarding as a hospital-wide problem and not an "ED problem," and they have elected to convert multiple sites around the hospital including pre-operative areas into boarding areas during hospital overcapacity instead of the ED, assigning ED to be the location of boarders as a true last resort.

3.2.2. Dependent Variables The primary outcomes of interest are both measures of efficiency and effectiveness in the ED. The first is patient length of stay, which we measure in two distinct ways. First, we measure the time from patient arrival until the attending physician completes care and determines disposition, capturing the duration until a decision is made to admit, discharge, or transfer the patient (Feizi et al. 2023). This metric specifically excludes delays related to inpatient bed availability, providing a clearer measure of ED operational efficiency. Second, we measure the total time a patient spends in the ED from arrival until physical departure (Lim et al. 2024). For admitted patients, this total time includes the duration until transfer to an "inpatient" bed, whether in the main hospital or in designated ED areas converted for inpatient use, encompassing both boarding time and discharge processing (Feizi et al. 2023). Given the documented right-skewed nature of ED time metrics (Song et al. 2015), we log-transform both time measurements to achieve approximately normal distributions (Brown et al. 2005, Saghafian et al. (2024)), meeting the assumptions required for our regression analyses.

Beyond time-based metrics, we examine resource utilization through the number of distinct imaging tests ordered during each ED encounter. This count variable helps us understand how batch ordering practices influence the overall diagnostic workload. To assess quality of care, we track whether patients return to the ED within 72 hours of their initial visit and require hospital admission (Lerman and Kobernick 1987). This binary indicator serves as a crucial quality metric, as returns within this timeframe often signal potential issues with initial treatment decisions, premature discharges, or missed diagnoses. All of these measures are widely used and validated in the emergency medicine literature, and are also of significant concern to our partner EDs. Furthermore, they allow us to evaluate ED performance across three critical dimensions: operational efficiency through time-based measurements, resource utilization via imaging orders, and care quality through return visit patterns. By examining these outcomes together, we can assess how batching behaviors and related and process changes might affect both the efficiency and effectiveness of care delivery.

3.3. Identification Strategy

Our empirical strategy closely follows the literature that relies on quasi-random assignment of agents to cases, often referred to as the "judges design." Papers in this literature typically exploit variation in the sentencing leniency of judges who work in the same court. Similarly, we explore batching variation across physicians who work in the same ED via a measure that we term "batch tendency." To measure physician batch tendency, we use rach physician's residualized leave-out average batch rate. We use this residualized measure of physician batch tendency because if certain physicians are more likely to work afternoon or weekend shifts (which Figure A1 shows to be the busiest shifts), the simple leave-out mean will be biased. The use of a residualized measure of physician batch tendency accounts for this kind of potential selection. This measure is derived from two steps following a similar approach used in some other applications (see, e.g., Doyle et al. (2015), Dobbie et al. (2018), and Eichmeyer and Zhang (2022)). First, we obtain residuals from a regression model, which includes all ED encounters in our sample period:

$$Batched_{i,t} = \alpha_0 + \alpha_{ym} + \alpha_{dt} + \beta \mathbf{X_{i,t}} + \varepsilon_{i,t}, \tag{1}$$

where $Batched_{i,t}$ is a dummy variable equal to one if patient i had their imaging tests batch ordered on an encounter that took place on date t. Fixed effects include year-month fixed effects, α_{ym}, to control for time and seasonal variation in batching, hospital-specific policies (e.g., initiatives to eliminate excess testing during a flu season) or seasonality in ED visits. We also control for "shift-level" variations that include both physician scheduling and patient arrival with day of week-time of day fixed effects, α_{dt}[2]. A vector of patient characteristics, $\mathbf{X_{i,t}}$, including chief complaint by ESI, vital signs, age, race, and sex was included to increase precision. As stated in 3.1, these controls are more than what is required for our quasi-random assignment assumption. Under the assumption that we have captured the observables under which quasi-random assignment occurs in the ED, the unexplained variation— the physician's contribution— resides in the error term, $\varepsilon_{i,t}$.

In step two, the tendency measure for patient i seen by physician j is computed as the average residual across all other patients seen by the physician during the study period:

$$BatchTendency^{phys}_{i,j} = \frac{1}{N_{-i,j}} \sum_{i' \in \{\mathbb{J} \setminus i\}} \hat{\varepsilon}_{i'} \tag{2}$$

where $\hat{\varepsilon}_{i'} = \hat{Batch}_{i'} - Batch_{i'}$ is the residual from Eq. (1); $\mathbb{J}$ is the set of all ED encounters treated by physician j; and $N_{-i,j} = |\{\mathbb{J} \setminus i\}|$, the number of cases that physician has seen that year, excluding patient i. This leave-out mean eliminates the mechanical bias that stems from patient i's own case entering into the instrument. The measure is interpreted as the average (leave-out) batch rate of patient i's physician, relative to other physicians in that hospital-year-month, hospital-day of week-time of day.

Figure 2 verifies that while the decision to batch depends on patient characteristics, our measure—batch tendency—is fairly exogenous and independent of patient characteristics. The left column of Figure 2 uses a linear probability model to test whether encounter and patient, ED, and physician characteristcs are predictive of the batching decision. We control for shift-level fixed effects and cluster standard errors at the physician level. We find that patients with

---

[2] Day of week takes on seven values: Sunday, Monday, etc., and time of day are six mutually exclusive four-hour bins: 8am–12 pm, 12pm–4 pm, etc.

Falls/Assaults/Trauma complaints are 16.2 percentage points more likely to be batched compared to similar patients under similar ED capacity. The right column assesses whether these features are predictive of our batch tendency measure using an identical specification. We find evidence that physicians of differing tendencies are assigned very similar patients. Put differently, the assignment of patients to physicians with high or low batch tendencies is fairly random, making the batch tendency measure an exogenous variable that could be exploited to study the impact of batching.

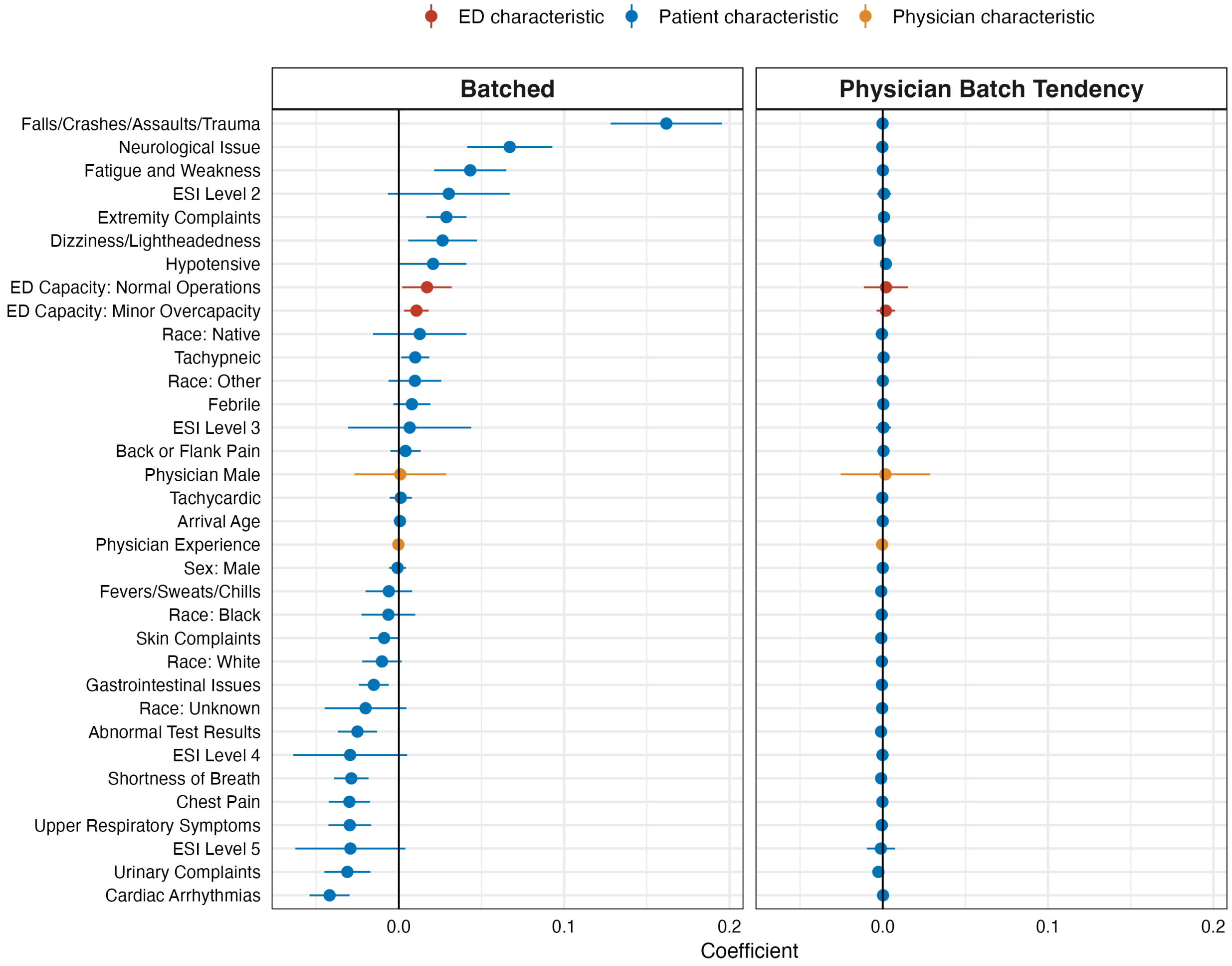

Figure 2 **Batch Tendency by Patient Characteristics**

Note: This figure plots a test for quasi-random assignment of patients to physicians in the Mayo Clinic ED. Residualization fixed effects include hospital-year-month, hospital-day of week-time of day. Robust standard errors are clustered at the physician level.

We further document that (a) there is significant variation in this measure, and (b) the measure is highly predictive of the decision to batch. Figure 3 shows the distribution of physician batch tendency and the relationship between batch tendency and batching, where the relationship is illustrated via local linear regression of batching against physician tendency. As can be seen, the probability of batching is monotonically, and approximately linearly, increasing in our tendency measure (see Table 2 for more formal results).

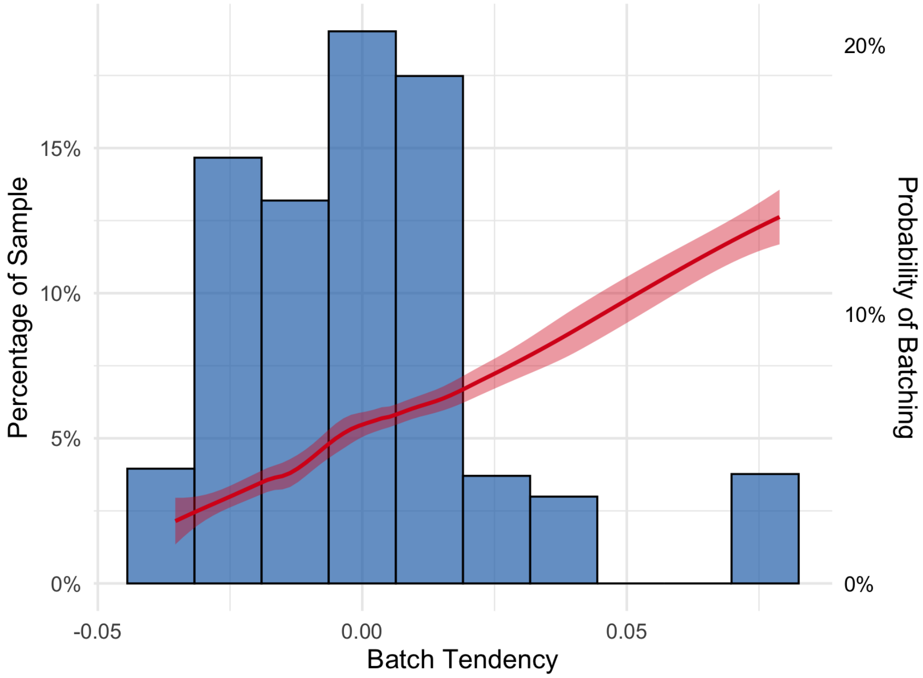

Figure 3 **Distribution and First Stage of Instrument**

Note: This figure plots the histogram of physician batch tendency along the x-axis and the left y-axis for all patient encounters. A local-linear regression of the fitted probability of batching on batch tendency after residualizing (see text for baseline fixed effects and controls in residualization) is overlaid and displayed on the right y-axis. Ninety-five percent confidence bands are also shown.

Put together, we observe that batch tendency likely satisfies the conditions required for a valid IV. In the next section, we formalize our IV analysis questions, and more formally establish the validity of our IV.

3.3.1. IV Analysis To estimate the reduced-form effects of being treated by a batch-preferring physician (batcher), we estimate the following equation:

$$Y_i = \mu_0 + \mu_1 \, BatchTendency_{i,j}^{phys} + \gamma \mathbf{X_i} + \nu_i \tag{3}$$

To study the effects of test batching on outcomes, Y_i, we estimate the following Two-Stage Least Squares (2SLS) equations:

$$\begin{cases} \text{First Stage:} & Batched_i = \delta_0 + \delta_1\, BatchTendency_{i,j}^{phys} + \delta_{\mathbf{2}}\mathbf{X_i} + \nu_i \\ \text{Second Stage:} & Y_i = \beta_0 + \beta_1 \hat{Batched_i} + \theta\mathbf{X_i} + \varepsilon_i \end{cases} \tag{4}$$

In all cases, Y_i represents our main outcomes of interest, and $\mathbf{X_i}$ includes the same covariates as in Eq. 1 and additional controls for physician experience, physician sex, and ED capacity level (ED capacity guidelines for Mayo Clinic are in Appendix Table A1. The variable $Batched_i$ may suffer from potential endogeneity concerns; for example, injury severity may be unobserved and correlated with the need to run multiple tests, length of stay, and return with admission likelihood. Hence, we instrument $Batched_i$ with the assigned physician j's underlying tendency to batch, $BatchTendency_{i,j}^{phys}$. We cluster robust standard errors at the physician level to account for the assignment process of patients to physicians.

Table 2 presents formal first-stage results from Eq. (4). Column 1 of Table 2 presents the mean crude batch rate. Column 2 begins by reporting results only with year-month and day of week-time of day fixed effects. Column 3 adds our baseline patient and hospital condition controls. Consistent with Figure 4, we find that our residualized physician instrument is highly predictive of whether a patient will have their imaging tests batch ordered. Including controls in column 3 does not change the magnitude of the estimated first-stage effect, consistent with the quasi-randomness of patient to physicians with different batching tendencies. Furthermore, we observe that batch tendency measure is fairly predictive of the batching decision, and the IV is not weak (F values are 335.8 and 536 for columns 2 and 3, respectively). For example, with all controls (column 3), our results show that a patient assigned to a physician that is at the 90th percentile of batch tendency (0.034) relative to a physician at the 10th percentile (-0.028) is 11.3 percentage points more likely to have their tests batched. The coefficient is greater than one because all emergency visits are used to construct the tendency instrument, while the first stage is calculated using the baseline sample only, which excludes the rare and rarely batched complaints.

3.3.2. Identifying Assumptions The reduced-form approach delivers an unbiased estimate of the causal effect of being treated by a higher tendency to batch physician, since assignment of patients to ED physicians is random, conditional on seasonality and shift ("conditional independence"). The residualization in Eq. (1) allows for further statistical precision in measuring physician tendency to batch.

Our instrumental variable approach, which aims to recover the causal effect of having diagnostic tests batch ordered, relies on three additional assumptions: relevance, exclusion, and monotonicity. We reported a strong first stage (i.e., relevance) at the end of the previous Section. The exclusion restriction requires that the instrument must influence the outcome of interest only through its effect on test batching. This is perhaps our strongest assumption and is at its core, untestable.

Table 2 First-Stage Results: Batch Tendency and Batching

	Sample Mean	Batch Tendency	
	(1)	(2)	(3)
Batched	0.134 (0.341)	1.786*** (0.097)	1.789*** (0.077)
Controls			
Shift-level Fixed Effects	—	Yes	Yes
Baseline Controls	—	No	Yes
Adj. R^2	—	0.023	0.064
F-stat	—	335.8	536.2
Observations	11,404	11,404	11,404

Notes: This table reports first-stage results for the regression of batch tendency on the likelihood of batching. Column 1 reports the sample mean and standard deviation of the dependent variable. Column 2 includes fixed effects for year-month and day of week-time of day. Column 3 adds baseline controls, including patient characteristics (age, vital signs like tachycardia, tachypnea, febrile, hypotensive status), hospital characteristics (capacity level), and physician characteristics (physician sex and physician experience). Robust standard errors are clustered at the physician level.

$^{***}p < 0.01$.

Nevertheless, we take this assumption seriously and perform a placebo check in Appendix D1 as well as various robustness checks in Section 4.8.

However, several features of the ED setting suggest that such violation may likely only have a small impact and may be less concerning than in other health care settings. First, unlike in primary care settings, where the patient and primary care provider have many repeated encounters, the scope of what the emergency physician can do to impact medium-term outcomes is limited and well-observed by the researcher. Second, any violation of the exclusion restriction needs to directly affect the specific outcome of interest. The channel by which ED physicians can influence the outcomes of interest are likely through testing and diagnosis. To the extent that the exclusion restriction is violated, our reduced form estimates can be interpreted as the causal impact of being assigned to a more or less likely to batch physician. These reduced form results are available in Table 3.

Finally, the monotonicity assumption is necessary for interpreting the coefficient estimates obtained from the IV approach as LATEs if there are heterogeneous treatment effects. It requires that any patient who is batched by a low batch tendency physician would also be batched by a high batch tendency physician. The literature leveraging the judges design typically performs two informal tests for its implications. The first one provides that the first stage should be weakly positive for all subsamples (Dobbie et al. (2018)). The second implication asserts that the instrument constructed by leaving out a particular subsample has predictive power over that same left-out subsample (Bhuller et al. (2020)). Appendix Figure D1 presents both of these tests in for various subsamples of interest. In the left panel, we find that our residualized measure of batch tendency is consistently positive and sizable in all subsamples, in line with the monotonicity assumption. In the right panel, we also find that our additional first-stage results are consistently same-signed and sizable across all subsamples. The coefficient magnitudes differ across subgroups because rates of batching differ.

3.3.3. More Details on Our LATE Estimates Our two-stage least squares estimates represent the LATE for patients whose likelihood of being batched changes as a result of their physician's batch tendency. To better understand this LATE, we characterize the number of compliers and their characteristics following the approach developed by Abadie and Gardeazabal (2003) and extended by Dahl et al. (2014b).

Specifically, compliers are defined as patients whose batching decision depends on whether their physician has the highest batch tendency ($\bar{z}$) or the lowest batch tendency ($\underline{z}$). Mathematically, the fraction of compliers (π_c) is given by:

$$\pi_c = P(Batched|Z = \bar{z}) - P(Batched|Z = \underline{z})$$

where Z represents the batch tendency of the physician. Using the first-stage regression model we predict the probabilities of being batched under the most lenient ($\bar{z}$) and strict ($\underline{z}$) physicians.

We find that approximately 13 percent of patients in our sample are "compliers," meaning they would have received batched tests if their physician had higher batch tendency but sequenced tests otherwise. In comparison, 5 percent of patients are "always takers," meaning they would receive batched tests regardless of the physician's batch tendency, and 82 percent are "never takers," meaning they would never receive batched tests regardless of the physician's batch tendency. In Appendix B, we provide more details on complier estimates and their characteristics.

4. Results and Discussion

4.1. Reduced-Form Results

In this sub-section, we start by exploring the causal influence of physician batch tendency on patient outcomes and resource utilization in the ED. We find statistically and operationally significant effects of assignment to a high batch tendency physician on every outcome except 72 hour return with admission (Table 3). Scaling our coeffecients by the difference in tendency going from the lowest decile to highest decile in physician tendency—equal to 6.3 percentage point increase—for interpretability, we find that assignment to a physician in the top batching decile (relative to one in the bottom decile) is associated with an 8.1% increase in LOS, an 8% increase in time to disposition,

Table 3 Reduced Form: Batch Tendency and Patient Outcomes

	Dependent variable			
	Log time to disposition (1)	Log LOS (2)	Number of distinct imaging tests (3)	72hr return with admission (4)
Batch tendency	1.526^{*}	1.589^{*}	2.561^{***}	-0.0222
	(0.6825)	(0.5811)	(0.2614)	(0.0364)
Mean dependent variable				
Time FE	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes
Adj R^2	0.152	0.104	0.236	0.011
Observations	11,361	11,361	11,361	11,361

Notes: This table reports the estimated coefficients of a reduced-form regression of our main outcomes on physician batch tendency. The dependent variables are log time to disposition, log LOS, number of distinct imaging tests, and 72-hour return with admission. Robust standard errors are clustered at the physician level.

$^{*}p < 0.05$, $^{***}p < 0.001$

and an additional 14.5 imaging tests ordered per 100 patient encounters. These findings highlight the substantial role ED physicians can play in putting patients on a path towards longer LOS and increased resource utilization.

The fact that our main outcomes respond so strongly to physician batch tendency is suggestive of batching as the underlying mechanism behind the effects. However, physicians could differ in other dimensions of care—some observable and others not—which could be correlated with batch tendency. Appendix D provides an attempt to distinguish between and identify the mechanisms behind the observed reduced-form effects. This mediation analysis provides a crucial step on the path toward a well-identified IV analysis.

4.2. Instrumental Variables Estimation

We next examine the effects of batch ordering imaging tests using our IV strategy described above. We first analyze the effects of batching on primary ED operational outcomes, before examining its impacts on specific test ordering patterns and disposition decisions.

Panel A of Table 4 presents OLS and 2SLS estimates of batching's impact on key operational metrics. Column 1 reports the dependent variable means for patients who had their tests ordered sequentially. Columns 2 and 3 report OLS estimates with increasing controls to understand potential sources of bias. Columns 4 and 5 report two-stage least squares results using our physician batch tendency instrument, without and with baseline controls respectively.

Table 4 IV Results: Batching Tests and Patient Outcomes

	Sequenced mean	OLS results		2SLS results	
	(1)	(2)	(3)	(4)	(5)
Panel A. Primary Outcomes					
Log time to disposition	5.236	0.083^{***}	0.111^{***}	0.784^{*}	0.854^{*}
	(0.451)	(0.014)	(0.013)	(0.366)	(0.285)
Log LOS	5.487	0.102^{***}	0.128^{***}	0.771^{*}	0.858^{*}
	(0.493)	(0.014)	(0.014)	(0.494)	(0.352)
Number of distinct imaging tests	1.338	0.836^{***}	0.826^{***}	1.381^{***}	1.403^{***}
	(0.575)	(0.017)	(0.016)	(0.183)	(0.184)
72hr return with admission	0.012	-0.002	0.000	-0.011	-0.009
	(0.111)	(0.002)	(0.002)	(0.014)	(0.013)
Panel B. Test Types					
Ultrasound	0.173	0.073^{***}	0.127^{***}	0.196^{*}	0.169^{*}
	(0.378)	(0.015)	(0.013)	(0.095)	(0.071)
CT with contrast	0.188	0.024	0.058^{*}	0.104^{*}	$0.151^{\dagger}$
	(0.390)	(0.015)	(0.014)	(0.083)	(0.081)
CT without contrast	0.402	0.383^{***}	0.280^{***}	0.118	0.148
	(0.490)	(0.015)	(0.013)	(0.134)	(0.127)
X-ray	0.575	0.355^{***}	0.360^{***}	0.963^{***}	0.959^{***}
	(0.494)	(0.012)	(0.013)	(0.185)	(0.219)
Panel C. Disposition					
Admission	0.288	$0.039^{\dagger}$	0.052^{*}	0.421^{***}	0.468^{***}
	(0.453)	(0.014)	(0.010)	(0.094)	(0.090)
Time FE	—	Yes	Yes	Yes	Yes
Baseline controls	—	No	Yes	No	Yes
Observations	11,361	11,361	11,361	11,361	11,361

Notes: This table reports the estimated coefficients of both an OLS and 2SLS regression of the effect of batching on patient outcomes. The OLS columns include no controls (column 2) and baseline controls (column 3). The 2SLS columns include no controls (column 4) and baseline controls (column 5). Standard errors are clustered at the physician level.

$^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The 2SLS estimates with full controls (column 5) indicate that batching substantially increases ED length of stay. The marginal batched patient experiences 85.8% increase in total ED length of stay and an 85.4% increase in time to disposition. These large effects suggest that in contrast to the fact that batching may seem efficient by initiating multiple diagnostic processes simultaneously, it extends patient average processing times significantly. Batching also leads to more intensive diagnostic testing. Specifically, the marginal batched patient receives 1.4 more distinct imaging tests, representing a 105% increase from the mean for sequentially tested patients. However, this increased testing intensity does not appear to impact quality of care, as measured by 72-hour returns with admission at the typical statistical significance levels. Notably, these findings suggest that the batching behavior in the ED results in additional tests that are delay-inducing butt not necessary value-adding. Put differently, they indicate that ordering diagnostics tests in a sequential manner offers an important benefit: the information obtained from prior tests reduce the need to order subsequent, non-value adding but delay-inducing tests (an "information gain" advantage).

Panel B of Table 4 examines how batching affects the utilization of specific imaging modalities. The two-stage least squares estimates reveal that batching leads to significant increases across all imaging types, with particularly large effects for X-rays. With full controls, batched patients are 95.9 percentage points more likely to receive an X-ray, 16.9 percentage points more likely to receive an ultrasound, 15.1 percentage points more likely to receive a contrast CT, and 14.8 percentage points more likely to receive a non-contrast CT. These substantial increases across all modalities suggest that physicians who batch tend to order more comprehensive imaging workups (compared to their non-batching peers) rather than substituting between different types of imaging.

Panel C of Table 4 presents estimates of batching's impact on ED disposition decisions. The two-stage least squares results indicate that batching substantially alters clinical decision-making patterns. With full controls, batching leads to a 46.8 percentage point increase in admission probability and a corresponding 40.1 percentage point decrease in discharge probability. These large, opposing effects suggest that batching pushes physicians toward more conservative disposition decisions. This also means that the impact of batching in the EDs likely spillovers to other parts of hospitals (e.g., inpatient units), increasing their patient volumes. Increased levels of patient volume, in turn, are known to induce behaviors such as speed-up which can have negative impact on other quality of care metrics such as 30-day mortality (Song and Saghafian 2019).

Taken together, these results paint a complex picture of batching's operational implications. While batching is associated with more comprehensive diagnostic workups, it also leads to significantly longer processing times and more conservative disposition decisions. The similar magnitudes between the OLS and two-stage least squares estimates suggest that selection bias may play a smaller role than anticipated in driving these relationships. However, the substantially larger standard errors in the two-stage least squares specifications indicate considerable heterogeneity in batching's effects across patients.

4.3. Potential Mechanism: Mediation Analysis

To better understand the mechanisms through which batching in the ED affects key operational outcomes— length of stay (LOS) and time to disposition— we investigated the mediating roles of imaging volume (number of tests) and admission decisions. Separate analyses were conducted for each operational outcome to ensure a comprehensive assessment of these pathways. These variables were selected as potential mediators for two reasons: (a) our prior analyses demonstrate that batching significantly increases both the number of tests ordered and the likelihood of admission, and (b) there is a strong theoretical basis to believe that these variables contribute to longer LOS and time to disposition. Increased imaging volume can delay patient processing due to the time required to conduct and interpret diagnostic tests. Similarly, resource utilization and diagnostic intensity are known to influence admission decisions, which directly impact ED throughput and LOS (Hodgson et al. 2018).

We formalized our hypothesized mechanisms using structural equation modeling (SEM) and the underlying DAG depicted in Figure 4. This approach allows us to estimate both the direct effects of batching on each operational outcome and the indirect effects mediated by imaging volume and/or admission decisions. Importantly, to control for potential confounding due to time and patient complexity, we first residualized all variables of interest by regressing them on fixed effects, including day of the week, month, and chief complaint severity. This residualization ensures that our estimates reflect associations net of these confounders, enabling a more precise exploration of the relationships among batching, mediators, and outcomes. For LOS, we included both imaging volume and admission decisions as mediators in the SEM. For time to disposition, however, admission decisions were excluded as a mediator because the disposition decision itself marks the endpoint of the outcome.

The results of the mediation analysis, detailed in Appendix Tables C1 and C2, shed light on a plausible mechanism by which batching impacts LOS and time to disposition. Our findings suggest that the effects of batching are primarily mediated through increased diagnostic testing intensity. This aligns with the results in Table 4, which demonstrated that batching leads to increased imaging volume. This pathway accounts for substantial delays in LOS, as additional imaging requires time for completion, interpretation, and integration into clinical decision-making. For LOS, the indirect effect via imaging volume was estimated at 0.207 ($p < 0.001$), indicating a substantial contribution to the total delay observed. Similarly, for time to disposition, the indirect effect via imaging was 0.085 ($p < 0.001$), underscoring the central role of diagnostic testing intensity in extending processing times. In contrast, the direct effects of batching on both LOS and time to disposition were small and statistically insignificant, as shown in Appendix Tables C1 and C2. This suggests that the delays attributed to batching are driven not by the act of batching itself, but rather by the subsequent increase in diagnostic testing activity that batching involves.

For LOS, the role of admission decisions adds complexity to this narrative. Affirming the results in Table 4, the SEM analysis (Table C1) revealed that batching was associated with a significant increase in admission likliehood ($b_2 = 0.327$, $p < 0.001$). Suggesting that increased imaging

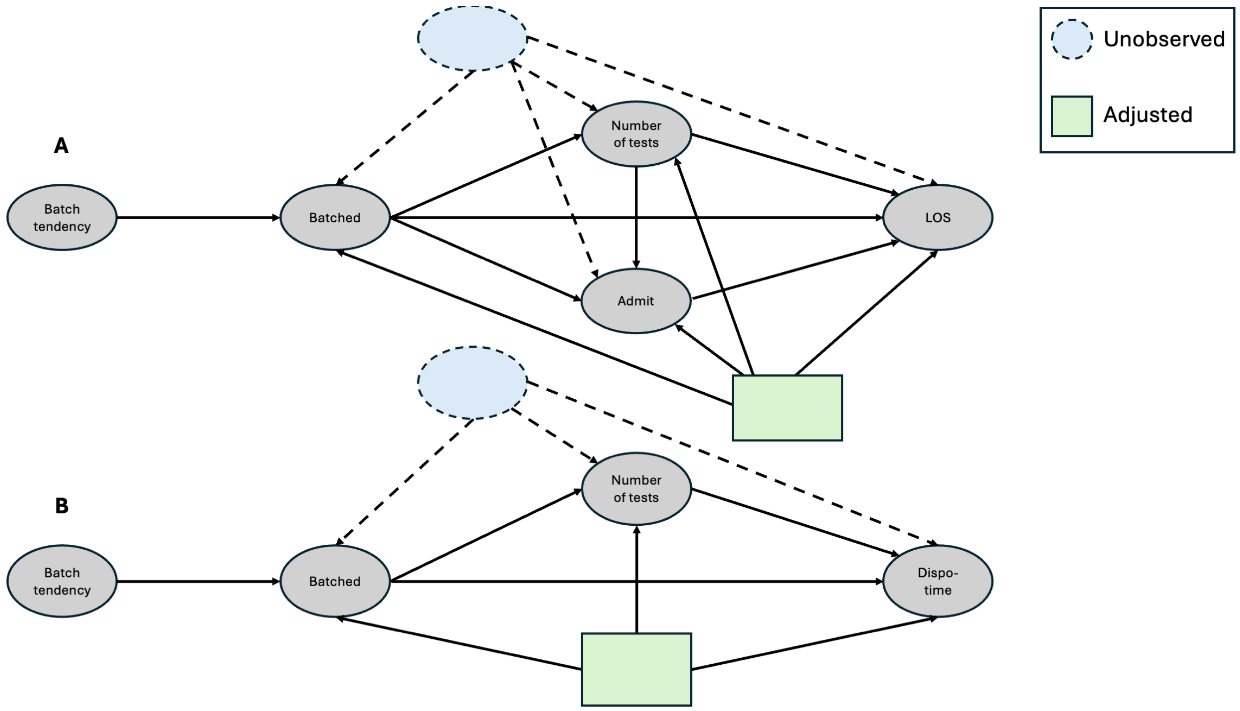

Figure 4 **Directed Acyclic Graph (DAG) of the Mediation Analysis**

may contribute to more conservative disposition decisions, potentially due to incidental findings or heightened diagnostic uncertainty, which necessitate further inpatient care. This pathway contributed additional delays to LOS, as coordinating patient transfers from the ED to inpatient units involves logistical challenges and further time consumption often accrued because of lack of bed availability in such units.. The total indirect effect for LOS, combining the pathways through imaging tests and admissions, was estimated at 0.256 ($p < 0.001$), while the direct effect of batching remained insignificant ($c' = 0.031$, $p = 0.662$). This consistency between the 2SLS and SEM results reinforces the notion that diagnostic intensity and conservative clinical decision-making are the key drivers of increased LOS in batched patients.

For time to disposition, where admission is not a mediator, the indirect effect through imaging tests (0.085, $p < 0.001$) remains the predominant pathway, with the direct effect of batching again nonsignificant ($c' = -0.100$, $p = 0.216$). This further highlights that the perceived efficiency gains from batching are outweighed by the operational burdens associated with increased diagnostic testing.

4.4. Heterogeneous Effects: ED Capacity Status

Given that ED capacity constraints significantly influence operational decisions, we next examine whether batching's effects vary across different capacity levels. Following The Mayo Clinic ED's internal guidelines[3] (Table A1), we categorize each encounter into to three categorizes: occurring during normal operations, minor overcapacity, or major overcapacity. While in all of our primary analysis we controlled for capacity status, stratifying by this allows us to understand whether the effectiveness of batching as a workflow strategy varies with ED congestion.

Table 5 Heterogeneous Effects of Batching by ED Capacity Status

	Normal Operations (1)	Minor Overcapacity (2)	Major Overcapacity (3)
Log LOS	0.890**	0.888*	0.534
	(0.290)	(0.354)	(0.602)
Log Time to Disposition	0.812*	0.895*	0.742
	(0.341)	(0.404)	(0.567)
Number of Distinct Imaging Tests	1.355***	1.474**	1.160**
	(0.088)	(0.398)	(0.357)
72hr Return with Admission	-0.023	-0.007	0.067
	(0.012)	(0.032)	(0.067)
Overall Batch Mean	0.158	0.133	0.099
Time FE	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes
Observations	3,742	5,846	1,816

Notes: This table reports two-stage least squares estimates of the impact of batching across different ED capacity levels. ED capacity levels are defined according to internal guidelines as described in Appendix Table A1. All specifications include time fixed effects and baseline controls. Robust standard errors clustered at the physician level are reported in parentheses.

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

The results in Table 5 show that batching's impact varies significantly across ED congestion levels, with important implications for operational efficiency. During normal operations, batching is most frequent (15.8% of encounters) and leads to substantial increases in both LOS and time to disposition, by 143% and 125%, respectively. Additionally, batching is associated with an increase of 1.36 imaging tests per patient. These findings suggest that when providers have greater discretion, batching is used more liberally, contributing to inefficiencies by increasing imaging volume and delaying result turnaround.

---

[3] The ED of our other partner hospital (MGH) does not follow these guidelines, and hence, we did not include these analyses in studying the effects of batching at MGH (see Section 4.6 for our analysis using MGH data).

As the ED enters minor overcapacity, batching frequency declines slightly to 13.3%, coinciding with the activation of workflow interventions like the Rapid Medical Assessment (RMA) process (Table A1). However, the negative effects of batching remain pronounced, with LOS increasing by 143% and time to disposition by 144%, while imaging volume rises even further (+1.47 tests per patient). This suggests that despite operational adaptations, providers continue using batching in a way that exacerbates inefficiencies, likely because these interventions do not directly constrain their ordering behavior.

A notable shift occurs under major overcapacity, where batching drops to 9.9% of encounters, aligning with stricter constraints such as potential ED diversion and expedited bed assignments. Here, the negative effects of batching are markedly reduced, with LOS and time to disposition increases are non-significant. Imaging increases as a result of batching also declines to 1.16 additional tests per patient. These results suggest that under extreme congestion, providers use batching more selectively, likely reserving it for clinically necessary cases rather than as a convenience.

These findings highlight how batching's impact depends on the ED's operational state. When providers have more discretion, batching appears inefficient, driving up imaging volume and delays. However, under severe congestion, batching becomes less frequent and less detrimental, suggesting a more strategic application. From a policy perspective, interventions to curb batching should focus on normal and minor overcapacity settings, where it contributes most to inefficiencies, while recognizing its potential necessity in extreme congestion.

4.5. Determinants of Image Batching

To investigate the drivers of batching and image ordering behavior, we examine the relationship between physician characteristics and ED crowding on the likelihood of batched testing. We estimate the following regression models:

$$Y_{i,j} = \beta_0 + \beta_1 \mathbf{MD_j} + \gamma Capacity + \alpha \mathbf{X_i} + \epsilon_i \tag{5}$$

Where $Y_{i,j}$ represents our outcome of interest: Batched, a binary measure of whether physician j batched tests for patient i, and the number of images ordered for patient i by physician j. $\mathbf{MD_j}$ is a vector of physician characteristics, including years since residency graduation, whether the physician is male, and the number of hours they are into their shift. *Capacity* is the current capacity level of the ED, defined in Appendix Table A1. $\mathbf{X_i}$ is the vectors of patient covariates described in the previous section and in Figure 2. We cluster robust standard errors at the physician level. Table 6 presents the results of the regression models.

We find that physician experience has minimal impact on batching behavior but is negatively associated with the total number of imaging tests ordered (-0.003, $p<0.05$), suggesting more experienced physicians tend to be more selective in their test ordering. While physician gender shows

Table 6 Determinants of Test Ordering Behavior

	Batched (1)	Number of Imaging Tests (2)
Panel A. Physician Characteristics		
Physician Experience	-0.0002	-0.003
	(0.001)	(0.002)
Physician Male	0.002	0.003
	(0.029)	(0.043)
Hours into Shift	-0.004***	-0.0003
	(0.001)	(0.003)
Panel B. ED Conditions		
Capacity Level: Minor Overcapacity	-0.021*	-0.033**
	(0.010)	(0.016)
Capacity Level: Major Overcapacity	-0.053***	-0.071***
	(0.016)	(0.025)
Time FE	Yes	Yes
Baseline controls	Yes	Yes
Observations	11,404	11,404

Notes: This table reports OLS estimates of the relationship between physician/ED characteristics and test ordering behavior. All specifications control for patient characteristics and time fixed effects. Robust standard errors clustered at the physician level are reported in parentheses.

* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

no significant relationship with either batching or overall test volume, we find that the timing within a physician's shift significantly influences batching decisions. Specifically, for each additional hour into the physician's shift, the likelihood of batching tests decreases by 0.3 percentage points ($p<0.01$). Because the Mayo Clinic ED features very few handoffs, and physicians tend to stay with their patients until disposition, this decline in batching as shifts progress may reflect physicians adopting more conservative testing strategies to ensure they can complete their work before their shift ends.

ED capacity conditions also show interesting relationships with test ordering patterns. Under both minor and major overcapacity conditions, we observe small but highly significant decreases (-0.001, $p<0.001$) in the number of imaging tests ordered, and larger negative affects on the likelihood of batching. This selective reduction in testing during periods of high ED occupancy stands in contrast with previous literature suggesting that ED crowding prompts less discernment in the use of diagnostic resources (Pines 2009).

4.6. Generalizability of Results Across EDs

To assess the generalizability of our findings beyond the Mayo Clinic ED, we replicated our analysis using data from the MGH ED, one of the busiest emergency departments in the United States. The MGH dataset comprises 129,489 patient encounters from November 10, 2021, through December 10, 2022. This extensive dataset provides a robust sample to validate the external applicability of our results.

Unlike the Mayo Clinic ED, where arriving patients are randomly assigned to physicians through a rotational system, the MGH ED employs a different patient assignment mechanism. At MGH, patients are triaged into different care areas based on acuity and presenting complaints. Within these areas, patients are assigned to physicians based on availability rather than a random rotational system. We note that this assignment could induce bias into our estimates, which is why we do not use MGH in our primary analysis. While this assignment could induce bias into our estimates, we adjust our instrumental variable strategy to account for these differences by including additional covariates for care area assigmnent, acuity level, and presenting complaint. While this adjustment may not fully account for the differences in patient assignment mechanisms, it provides a more robust comparison of the effects of batching on patient outcomes across different hospital settings.

Table 7 Comparison of Effects Across Hospital Settings

	Sequenced Mean (SD) (1)	OLS (2)	2SLS (3)	Z-Statistic (4)
Log LOS	6.42	0.21^{***}	0.34	-0.907
	(0.847)	(0.010)	(0.450)	
Number of distinct imaging tests	1.34	0.780^{***}	1.835^{***}	1.23
	(0.563)	(0.009)	(0.320)	
72hr return with admission	0.0123	-0.0048^{***}	0.0158	0.617
	(0.110)	(0.001)	(0.038)	
Time FE	—	Yes	Yes	—
Baseline controls	—	Yes	Yes	—
Observations	—	42,085	42,085	—

Notes: This table reports OLS and two-stage least squares estimates from the MGH dataset. All specifications include time fixed effects, baseline controls, and care area fixed effects. Robust standard errors clustered at the physician level are reported in parentheses. Column 1 reports the mean and standard deviation for non-batched patients (sequenced). Column 4 reports the Z-statistic from a formal test comparing the 2SLS coefficient in Column 3 to the 2SLS coefficient in Column 5 of Table 4.

Significance levels: $^{***}p < 0.001$

After adjusting for institutional differences and using the same exclusion criteria as we used with Mayo, we find strong evidence that our key findings generalize to the MGH setting. The 2SLS results in Table \ref{tab:generalize suggest batching leads to a 44.3% increase in length of stay and approximately 1.8 additional imaging tests per patient. To formally assess whether the estimated effects differ significantly across the two ED settings, we conduct a Z-test comparing the 2SLS coefficients from MGH and Mayo by estimating:

$$Z = \frac{\hat{\beta}\text{MGH} - \hat{\beta}\text{Mayo}}{\sqrt{SE^2_{\text{MGH}} + SE^2_{\text{Mayo}}}}$$

.

As reported in Column 4, the Z-statistics for each outcome indicate no statistically significant differences in the estimated effects between the two settings. This suggests that the impact of batching is largely consistent across hospitals, despite differences in patient assignment mechanisms and operational structures. Overall, these results reinforce the external validity of our findings and provide further evidence that the observed effects of batching are not merely an artifact of a single institution's workflow but rather a systematic consequence of batching in high-volume emergency care settings.

4.7. Managerial Implications

Our findings have important implications for ED operations management. First, we show that batch ordering of imaging tests, while potentially appealing as a workflow efficiency strategy, can significantly increase both length of stay and resource utilization without corresponding improvements in patient outcomes. The substantial magnitude of these effects—an 85.8% increase in LOS for batched patients—suggests that ED managers should carefully evaluate policies around physician test ordering discretion. As our mediation analysis reveals, these delays arise primarily through increased diagnostic intensity rather than the batching process itself, indicating that sequential ordering may serve as a natural filter against unnecessary testing.

The significant variation we observe in batching behavior across physicians treating similar patients suggests an opportunity for standardization. ED managers might consider implementing decision support systems that encourage sequential ordering, particularly for conditions where the information gained from initial tests frequently eliminates the need for additional imaging. However, our heterogeneity analysis indicates that such policies should be flexible to ED conditions—the reduced impact of batching on test ordering during periods of major overcapacity suggests that different strategies may be optimal under varying capacity constraints. This would specifically the case, if for example ED physicians consciously or subconsciously avoid over-testing when they know resources are significantly limited.

The substantial cost implications of batch ordering—both in terms of operational efficiency and resource utilization—suggest that EDs could benefit from more structured approaches to test ordering. While preserving physician autonomy in clinical decision-making is crucial, our results indicate that completely unfettered discretion in test ordering timing may lead to suboptimal system performance. Simple interventions like providing feedback to physicians about their individual batching rates relative to peers, or implementing “nudge” systems that suggest sequential ordering pathways, could help reduce unnecessary testing while maintaining quality of care.

Finally, our findings have implications beyond individual EDs, as increased admission rates from batch ordering create spillover effects throughout the hospital system. Hospital administrators should consider these downstream impacts when developing imaging protocols and resource allocation strategies. The fact that batching leads to both increased testing and higher admission rates suggests that policies aimed at reducing unnecessary batch ordering could yield benefits across multiple dimensions of hospital operations.

4.8. Robustness Checks

While our research design leverages the random assignment of patients to physicians to identify causal effects, several limitations warrant discussion. A primary concern is that physicians’ tendency to batch order tests may correlate with other unobserved practice patterns that affect our outcomes of interest. Although we found no significant associations between observable physician characteristics (such as experience, gender, or training) and batch ordering tendency, unobserved characteristics could influence both the tendency to batch and other aspects of patient care. For instance, physicians who tend to batch order tests might also have different approaches to patient assessment, documentation practices, or consultation patterns that independently affect length of stay and disposition decisions. Removing the potential impact of such unobserved factors might ultimately require running a fully randomized experiment. However, the consistency and magnitude of our findings across (a) both OLS and 2SLS specifications, and (b) two different hospitals with different practice settings provide strong evidence behind our findings. In particular, unobserved physician characteristics would need to be substantial to invalidate our core finding that batching increases both test utilization and processing times.

Nevertheless, the validity of our results largely depend on our identification strategy assumptions (Section 3.3). Thus, we performed several robustness checks (see Appendix D) to gain further confidence about the validity of our results. To probe the conditional independence assumption behind our identification strategy, we replaced the physician tendency instrument with a shift-level instrument that captures the average tendency across all physicians working during the same shift. To provide robustness checks to potential exclusion restriction violations, we adjusted for chief observable margins of care that may be correlated with tendency to batch, namely (a) decision to admit patients to the hospital, (b) propensity to order laboratory tests, and (c) propensity to order imaging tests. Additionally we provide a placebo test where we check whether the instrument has a positive first stage for a placebo outcome, namely the number of laboratory tests ordered. Furthermore, we investigate whether the reduced-form effects observed in Section 4.1 are due to differences in batch ordering rates across physicians or due to other provider differences correlated with batching tendency. We examine this by studying reduced-form effects among patients with conditions that rarely require multiple imaging tests, as a falsification check. Finally, to address monotonicity, we construct physician-complaint-specific leniency instruments, and we check whether our instrument has a positive first stage across key complaint, severity, capacity, and demographic related subsamples (Appendix Figure D1). Across all robustness checks and main outcomes, we found that magnitudes of our estimates remain fairly unchanged, and the implications of our results remain consistent. Our robustness checks, thus, give us further confidence about the validity of our main findings. We do, however, acknowledge that a randomized experiment might be ultimately needed to verify the results obtained from our observational data.

5. Conclusion

Although prior literature examines task ordering from the perspective of centralized protocols, frontline physicians often have discretion over how and when they order diagnostic tests, especially in high-pressure environments like EDs. In practice, either due to system design or individual choice, this delegation of decision-making results in physicians self-managing their test ordering strategies. Given the limited research on the operational implications of test ordering, little is known about how healthcare managers should guide these practices when such discretion exists. Understanding when and how physicians exercise this discretion informs decisions about system design, the extent to which discretion should be granted, how to encourage optimal use of discretion, and how to adjust policies to account for the behaviors of frontline clinicians.

We explore this underexamined area by analyzing the drivers and consequences of batch ordering imaging tests in the ED. Utilizing detailed operational data from two large U.S. EDs with random patient-physician assignment, we find that physicians are more likely to batch order imaging tests earlier in their shifts and when the ED is less crowded. This suggests that time pressure and occupancy levels significantly influence the decision to batch, consistent with the notion that physicians may adjust their ordering behavior based on their workload and the operational demands of the ED.

Our results indicate that batch ordering imaging tests leads to a significant increase in the number of imaging studies performed per patient encounter, confirming that batching contributes to higher resource utilization. However, we do not find evidence that batching reduces patient LOS in the ED or impacts the likelihood of a 72-hour return with admission. On the contrary, we find that batching greatly increases the time spent in the ED and the time to disposition, suggesting that large inefficiencies from initiating multiple diagnostic imaging processes simultaneously. This result aligns with previous research emphasizing the importance of diagnostic pathways in achieving optimal health outcomes and operational efficacy (Carpenter et al. 2015, Masic et al. (2008), Singh and Graber (2015)). This is consistent with the information gain advantage of sequential test ordering, where the results of one test may eliminate the need for another. Our findings suggest that while discretion allows physicians to tailor their test ordering to specific situations, it may also lead to practices that are resource-intensive without corresponding benefits to patients or the ED system.

Moreover, over-testing in EDs is not a benign phenomenon. It is associated with increased risks, including patient exposure to unnecessary radiation and the resultant psychological and physical burden from incidental findings (Müskens et al. 2022). Moreover, the economic implications are substantial, with the overuse of diagnostic tests contributing significantly to the escalating costs of healthcare (Atkinson and Saghafian 2023). As such, our results suggest the need to examine the practice of batching across different clinical conditions and in other clinical settings beyond the ED, and also consider it is consequences across a verity of metrics that affect hospital's publicly reported outcomes (Saghafian and Hopp 2020, Saghafian and Hopp (2019)).

Our study underscores the importance of developing evidence-based guidelines to inform physicians' test ordering strategies. By understanding how batching impacts on patient outcomes and ED operations, healthcare managers can design interventions to optimize test ordering practices. This may include providing decision support tools, adjusting policies to encourage sequential ordering when appropriate, or offering feedback to physicians on their ordering patterns and associated outcomes. By aligning physician test ordering behaviors more closely with patient needs (Atkinson and Saghafian 2023), EDs may enhance patient satisfaction and outcomes while improving operational efficiency.

References

Abadie A, Gardeazabal J (2003) The economic costs of conflict: A case study of the basque country. *American Economic Review* 93(1):113–132.

Atkinson MK, Saghafian S (2023) Who should see the patient? on deviations from preferred patient-provider assignments in hospitals. *Health Care Management Science* 26(2):165–199, URL `http://dx.doi.org/10.1007/s10729-022-09628-x`, epub 2023 Apr 27.

Baloescu C (2018) Diagnostic imaging in emergency medicine: How much is too much? *Annals of Emergency Medicine* 72(6):637–643.

Balogh E, Miller B, Ball J (2015) *Improving Diagnosis in Health Care* (Washington (DC): National Academies Press (US)).

Batt RJ, Terwiesch C (2016) Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science* 63(11):3531–3551, URL `http://dx.doi.org/10.1287/mnsc.2016.2543`.

Bendoly E (2011) Linking task conditions to physiology and judgment errors in rm systems. *Production and Operations Management* 20(6):860–876, URL `http://dx.doi.org/10.1111/j.1937-5956.2011.01230.x`.

Berry Jaeker JA, Tucker AL (2020) The value of process friction: The role of justification in reducing medical costs. *Journal of Operations Management* 66(1-2):12–34, URL `http://dx.doi.org/10.1002/joom.1076`.

Bhuller M, Dahl GB, Løken KV, Mogstad M (2020) Incarceration, recidivism, and employment. *Journal of Political Economy* 128(4):1269–1324.

Boudreau J, Hopp W, McClain JO, Thomas LJ (2003) On the interface between operations and human resources management. *Manufacturing & Service Operations Management* 5(3):179–202.

Brown L, Gans N, Mandelbaum A, Sakov A, Shen H, Zeltyn S, Zhao L (2005) Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American Statistical Association* 100(469):36–50, URL `http://dx.doi.org/10.1198/016214504000001808`.

Campbell D, Frei FX (2011) Market heterogeneity and local capacity decisions in services. *Manufacturing & Service Operations Management* 13(1):2–19, URL `http://dx.doi.org/10.1287/msom.1100.0312`.

Carpenter C, Raja A, Brown M (2015) Overtesting and the downstream consequences of overtreatment: Implications of ”preventing overdiagnosis” for emergency medicine. *Academic Emergency Medicine* 22(12):1484–1492, URL `http://dx.doi.org/10.1111/acem.12820`.

Chan DC (2018) The efficiency of slacking off: Evidence from the emergency department. *Econometrica* 86(3):997–1030, URL `http://dx.doi.org/10.3982/ECTA13565`.

Chisholm CD, Collison EK, Nelson DR, Cordell WH (2000) Emergency department workplace interruptions: are emergency physicians ”interrupt-driven” and ”multitasking”? *Academic Emergency Medicine* 7(11):1239–1243, URL `http://dx.doi.org/10.1111/j.1553-2712.2000.tb00469.x`.

Chisholm CD, Dornfeld AM, Nelson DR, Cordell WH (2001) Work interrupted: A comparison of workplace interruptions in emergency departments and primary care offices. *Annals of Emergency Medicine* 38(2):146–151, URL `http://dx.doi.org/10.1067/mem.2001.115440`.

Cournane S, Conway R, Creagh D, Byrne D, Sheehy N, Silke B (2016) Radiology imaging delays as independent predictors of length of hospital stay for emergency medical admissions. *Clinical Radiology* 71(9):912–918, URL `http://dx.doi.org/10.1016/j.crad.2016.03.023`.

Coussens S, Ly DP (2024) Variation in emergency department physician admitting practices and subsequent mortality. *JAMA Internal Medicine* URL `http://dx.doi.org/10.1001/jamainternmed.2024.6925`, published online.

Dahl GB, Kostøl AR, Mogstad M (2014a) Family welfare cultures. *The Quarterly Journal of Economics* 129(4):1711–1752, URL `http://dx.doi.org/10.1093/qje/qju019`.

Dahl GB, Løken KV, Mogstad M (2014b) Peer effects in program participation. *American Economic Review* 104(7):2049–2074.

Daniels M, Schroeder SA (1977) Variation among physicians in use of laboratory tests ii. relation to clinical productivity and outcomes of care. *Medical Care* 15(6):482–487, URL `http://dx.doi.org/10.1097/00005650-197706000-00004`.

Deo S, Jain A (2019) Slow first, fast later: Temporal speed-up in service episodes of finite duration. *Production and Operations Management* 28(5):1061–1081, URL `http://dx.doi.org/10.1111/poms.12955`.

Dobbie W, Goldin J, Yang CS (2018) The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review* 108(2):201–240.

Doyle J, Graves J, Gruber J, Kleiner S (2015) Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy* 123(1):170–214, URL `http://dx.doi.org/10.1086/677756`.

Eichmeyer S, Zhang J (2022) Pathways into opioid dependence: Evidence from practice variation in emergency departments. *American Economic Journal: Applied Economics* 14(4):271–300, URL `http://dx.doi.org/10.1257/app.20210048`.

Feizi A, Carson A, Jaeker JB, Baker WE (2023) To batch or not to batch? impact of admission batching on emergency department boarding time and physician productivity. *Operations Research* 71(3):939–957, URL `http://dx.doi.org/10.1287/opre.2022.2381`.

Fowler JW, Mönch L (2022) A survey of scheduling with parallel batch (p-batch) processing. *European Journal of Operational Research* 298(1):1–24, ISSN 0377-2217, URL `http://dx.doi.org/10.1016/j.ejor.2021.06.012`.

Gowrisankaran G, Joiner K, Léger PT (2022) Physician practice style and healthcare costs: Evidence from emergency departments. *Management Science* 69(6):3202–3219, URL `http://dx.doi.org/10.1287/mnsc.2022.4470`.

Hodgson NR, Saghafian S, Mi L, Buras MR, Katz ED, Pines JM, Sanchez L, Silvers S, Maher SA, Traub SJ (2018) Are testers also admitters? comparing emergency physician resource utilization and admitting practices. *The American Journal of Emergency Medicine* 36(10):1865–1869, URL `http://dx.doi.org/10.1016/j.ajem.2018.07.045`.

Ibanez M, Toffel MW (2020) How scheduling can bias quality assessment: Evidence from food safety inspections. *Management Science* 66(6):2396–2416, URL `http://dx.doi.org/10.1287/mnsc.2019.3333`.

Ibanez MR, Clark JR, Huckman RS, Staats BR (2018) Discretionary task ordering: Queue management in radiological services. *Management Science* 64(9):4389–4407, URL `http://dx.doi.org/10.1287/mnsc.2017.2800`.

Jameson J, Saghafian S, Huckman R, Hodgson N (2024) Variation in batch ordering of imaging tests in the emergency department and the impact on care delivery. *Health Services Research* URL `http://dx.doi.org/10.1111/1475-6773.14406`, epub ahead of print.

Jessome R (2020) Improving patient flow in diagnostic imaging: a case report. *Journal of Medical Imaging and Radiation Sciences* 51(4):678–688, URL `http://dx.doi.org/10.1016/j.jmir.2020.08.014`, epub 2020 Sep 17. PMID: 32950432; PMCID: PMC7495148.

Juliusson G, Thorvaldsdottir B, Kristjansson JM, Hannesson P (2019) Diagnostic imaging trends in the emergency department: an extensive single-center experience. *Acta Radiologica Open* 8(7):2058460119860404, URL `http://dx.doi.org/10.1177/2058460119860404`.

KC DS (2013) Does multitasking improve performance? evidence from the emergency department. *Manufacturing & Service Operations Management* 16(2):168–183, URL `http://dx.doi.org/10.1287/msom.1120.0409`.

KC DS, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498, URL `http://dx.doi.org/10.1287/mnsc.1090.1048`.

Leppink J, Hanham J (2019) Mental processes in emergency medicine. *Instructional Design Principles for High-Stakes Problem-Solving Environments*, 55–62 (Springer).

Lerman B, Kobernick MS (1987) Return visits to the emergency department. *The Journal of Emergency Medicine* 5(5):359–362, URL `http://dx.doi.org/10.1016/0736-4679(87)90241-9`.

Lim JM, Song H, Yang JJ (2024) The spillover effects of capacity pooling in hospitals. *Management Science* 70(11):7692–7711, URL `http://dx.doi.org/10.1287/mnsc.2022.02202`.

Lyu H, Xu T, Brotman D, Mayer-Blackwell B, Cooper M, et al. (2017) Overtreatment in the united states. *PLOS ONE* 12(9):e0181970, URL `http://dx.doi.org/10.1371/journal.pone.0181970`.

Masic I, Miokovic M, Muhamedagic B (2008) Evidence based medicine-new approaches and challenges. *Acta Informatica Medica* 16(4):219, URL `http://dx.doi.org/10.5455/aim.2008.16.219-225`.

Miller RA (1994) Medical diagnostic decision support systems–past, present, and future: a threaded bibliography and brief commentary. *Journal of the American Medical Informatics Association* 1(1):8–27, URL `http://dx.doi.org/10.1136/jamia.1994.95236141`.

Mills AM, Raja AS, Marin JR (2015) Optimizing diagnostic imaging in the emergency department. *Academic Emergency Medicine* 22(5):625–631, URL `http://dx.doi.org/10.1111/acem.12640`, epub 2015 Mar 2.

Müskens J, Kool R, van Dulmen S, Westert G (2022) Overuse of diagnostic testing in healthcare: a systematic review. *BMJ Quality & Safety* 31(1):54–63, URL `http://dx.doi.org/10.1136/bmjqs-2020-012576`.

Perotte R, Lewin G, Tambe U, et al. (2018) Improving emergency department flow: Reducing turnaround time for emergent ct scans. *AMIA Annual Symposium Proceedings*, 897–906.

Pines JM (2009) Trends in the rates of radiography use and important diagnoses in emergency department patients with abdominal pain. *Medical Care* 47(7):782–786, URL `http://dx.doi.org/10.1097/MLR.0b013e31819432a0`.

Pines JM (2017) What cognitive psychology tells us about emergency department physician decision-making and how to improve it. *Academic Emergency Medicine* 24(1):117–119, URL `http://dx.doi.org/10.1111/acem.13061`.

Poyiadji N, Beauchamp Nr, Myers DT, Krupp S, Griffith B (2023) Diagnostic imaging utilization in the emergency department: Recent trends in volume and radiology work relative value units. *Journal of the American College of Radiology* 20(12):1207–1214, URL `http://dx.doi.org/10.1016/j.jacr.2023.06.033`, epub 2023 Aug 4.

Raja AS, Ip IK, Sodickson AD, et al. (2014) Radiology utilization in the emergency department: Trends of the past 2 decades. *AJR American Journal of Roentgenology* 203:355–360, URL `http://dx.doi.org/10.2214/AJR.13.11892`.

Rogg JG, Huckman R, Lev M, Raja A, Chang Y, White BA (2017) Describing wait time bottlenecks for ed patients undergoing head ct. *The American Journal of Emergency Medicine* 35(10):1510–1513, ISSN 0735-6757, URL `http://dx.doi.org/10.1016/j.ajem.2017.04.059`.

Saghafian S, Austin G, Traub SJ (2015) Operations research/management contributions to emergency department patient flow optimization: Review and research prospects. *IIE Transactions on Healthcare Systems Engineering* 5(2):101–123.

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

Saghafian S, Hopp WJ (2019) The role of quality transparency in health care: Challenges and potential solutions. *NAM Perspectives* 2019:10.31478/201911a, URL `http://dx.doi.org/10.31478/201911a`.

Saghafian S, Hopp WJ (2020) Can public reporting cure healthcare? the role of quality transparency in improving patient–provider alignment. *Operations Research* 68(1):71–92, URL `http://dx.doi.org/10.1287/opre.2019.1901`.

Saghafian S, Kilinc D, Traub SJ (2024) Dynamic assignment of patients to primary and secondary inpatient units: Is patience a virtue? Grosskopf S, Valdmanis V, Zelenyuk V, eds., *The Cambridge Handbook of Healthcare: Productivity, Efficiency, Effectiveness*, 612–656 (Cambridge University Press).

Singh H, Graber M (2015) Improving diagnosis in health care—the next imperative for patient safety. *New England Journal of Medicine* 373(26):2493–2495, URL `http://dx.doi.org/10.1056/NEJMp1512241`.

Skaugset LM, Farrell S, Carney M, Wolff M, Santen SA, Perry M, Cico SJ (2016) Can you multitask? evidence and limitations of task switching and multitasking in emergency medicine. *Annals of Emergency Medicine* 68(2):189–195, URL `http://dx.doi.org/10.1016/j.annemergmed.2015.10.016`.

Smith-Bindman R, Kwan ML, Marlow EC, et al (2019) Trends in use of medical imaging in us health care systems and in ontario, canada, 2000-2016. *JAMA* 322(9):843–856, URL `http://dx.doi.org/10.1001/jama.2019.11456`.

Smulowitz PB, O'Malley AJ, McWilliams JM, Zaborski L, Landon BE (2021) Variation in rates of hospital admission from the emergency department among medicare patients at the regional, hospital, and physician levels. *Annals of Emergency Medicine* 78(4):474–483, ISSN 0196-0644, URL `http://dx.doi.org/10.1016/j.annemergmed.2021.03.020`.

Solomon DH, Hashimoto H, Daltroy L, Liang MH (1998) Techniques to improve physicians' use of diagnostic tests: a new conceptual framework. *JAMA* 280(23):2020–2027, URL `http://dx.doi.org/10.1001/jama.280.23.2020`.

Soltani M, Batt RJ, Bavafa H, Patterson BW (2022) Does what happens in the ed stay in the ed? the effects of emergency department physician workload on post-ed care use. *Manufacturing & Service Operations Management* 24(6):3079–3098, URL `http://dx.doi.org/10.1287/msom.2022.1110`, epub 2022 May 2.

Song H, Tucker AL, Murrell KL (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science* 61(12):3032–3053, URL `http://dx.doi.org/10.1287/mnsc.2014.2091`.

Song H, Tucker AL, Murrell KL, Vinson DR (2017) Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Science* 64(6):2628–2649, URL `http://dx.doi.org/10.1287/mnsc.2016.2642`.

Song L, Saghafian S (2019) The spillover effects of hospital closures on the efficiency and quality of other hospitals. HKS Working Paper RWP19-006, Harvard Kennedy School, URL `http://dx.doi.org/10.2139/ssrn.3318609`.

Tamburrano A, Vallone D, Carrozza C, et al. (2020) Evaluation and cost estimation of laboratory test overuse in 43 commonly ordered parameters through a computerized clinical decision support system (ccdss) in a large university hospital. *PLoS One* 15(8):e0237159, URL `http://dx.doi.org/10.1371/journal.pone.0237159`.

Traub S, Saghafian S, Judson K, et al. (2018) Interphysician differences in emergency department length of stay. *Journal of Emergency Medicine* 54(5):702–710.

Traub SJ, Stewart CF, Didehban R, Bartley AC, Saghafian S, Smith VD, Silvers SM, LeCheminant R, Lipinski CA (2016) Emergency department rotational patient assignment. *Annals of Emergency Medicine* 67(2):206–215, URL `http://dx.doi.org/10.1016/j.annemergmed.2015.07.008`, epub 2015 Oct 6. PMID: 26452721.

Valtchinov VI, Ip IK, Khorasani R, Schuur JD, Zurakowski D, Lee J, Raja AS (2019) Use of imaging in the emergency department: Do individual physicians contribute to variation? *AJR American Journal of Roentgenology* 213(3):637–643, URL `http://dx.doi.org/10.2214/AJR.18.21065`.

van Donselaar KH, Gaur V, van Woensel T, Broekmeulen RACM, Fransoo JC (2010) Ordering behavior in retail stores and implications for automated replenishment. *Management Science* 56(5):766–784, URL `http://dx.doi.org/10.1287/mnsc.1100.1146`.

Wennberg JE (1984) Dealing with medical practice variations: a proposal for action. *Health Affairs (Millwood)* 3(2):6–32, URL `http://dx.doi.org/10.1377/hlthaff.3.2.6`.