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 examine practice variation across physicians to uncover the operational impact of discretionary batch ordering of imaging tests. We find that quasi-random assignment of a patient to an ED physician who is a "batcher" (top decile) versus a "sequencer" (bottom decile) causally increases the patient's length of stay, time to disposition, and the number of imaging tests performed. Instrumental variable results show that discretionary batching increases length of stay by approximately 65%, increases imaging tests by 88%, and has no effect on 72-hour returns. Through a decomposition analysis, we find that approximately one-quarter of the LOS increase is attributable to increased imaging volume, with hospital admission as an additional contributing pathway, and the remainder consistent with increased coordination costs and decision complexity inherent to managing concurrent diagnostic workflows. Results suggest that discretionary batching may lead to clinical decision-making that introduces bottlenecks in patient flow. Conversely, standard practice, which preserves diagnostic flexibility by allowing information from initial tests to guide subsequent decisions, offers an "information gain" advantage over batch ordering: the information obtained from a prior test allows the elimination of the need to order some future tests. Put together, our findings indicate that discretionary batch ordering is not an efficient strategy for managing diagnostic imaging in emergency care, and interventions to reduce unnecessary batching may improve both operational performance and resource utilization.

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 its most significant operational bottlenecks (Rogg et al. 2017). Advanced diagnostic

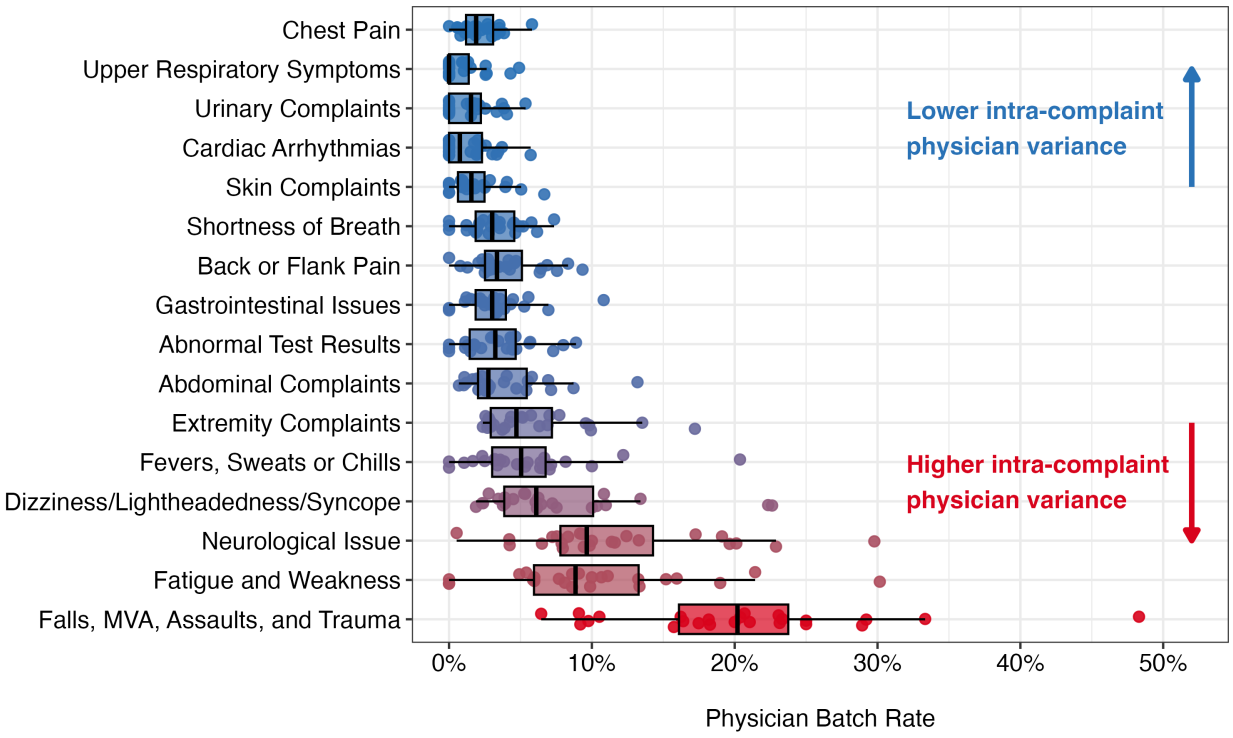
imaging has risen dramatically over the past two decades, transforming from a limited resource to a cornerstone of emergency care (Juliusson et al. 2019, Smith-Bindman et al. (2019)). However, this transformation has intensified operational challenges in EDs, where the complexity stems from multiple 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)).

Though critical for the diagnosis of many conditions, 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 radiation exposure, 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 diagnostic testing management is critical for patient outcomes and ED operations (Balogh et al. 2015). Physicians have considerable discretion over how they order these tests, and wide variation in diagnostic test ordering behavior has been well documented (Miller 1994, Solomon et al. (1998), Wennberg (1984), Daniels and Schroeder (1977)). However, relatively little research has considered how to manage test ordering strategies when this discretion exists.

We consider an ED physician's decision to order imaging tests for a patient as an optimization problem, where the physician must balance the tradeoffs between the advantages of ordering multiple tests simultaneously (batching) 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 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 physicians sequentially ordering one test at a time, reviewing the results, and then deciding whether to order additional tests based on the information obtained from the previous test. While this standard practice may serve as a natural filter to prevent unnecessary testing, it can also result in longer delays than batching when multiple tests are ultimately needed. This tension between potential efficiency gains and the risk of unnecessary testing is significant 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 with consideration of immediate operational implications and broader quality-of-care considerations (Feizi et al. 2023). This decision is inherently complex, requiring physicians

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 in which the physician batch-ordered imaging tests for a complaint by the number of encounters with that complaint.

to weigh time-sensitive diagnostic needs against resource constraints and the risk of ordering tests that may prove unnecessary once earlier results become available.

This paper explores the causal effects of batch ordering advanced imaging tests on operational performance and patient outcomes in the ED. Specifically, we focus on batch orders that include multiple types of imaging tests that cannot be run in a single scanning session. Using data from two leading U.S. hospitals—Mayo Clinic and Massachusetts General Hospital—we first provide evidence of significant variation in emergency physicians' batching behavior. Batching varies significantly among physicians working in the same ED, treating patients with the same complaint and severity (Figure 1). Using this variation, we next investigate whether being seen by a "batcher" or a "sequencer" physician has implications in terms of performance metrics such as length of stay (LOS), number of imaging studies, 72-hour rate of return, and disposition decision (i.e., being admitted to the hospital or discharged home post ED service). Finally, we shed light on the circumstances under which physicians are more likely to batch order imaging tests.

Generating causal evidence on batching presents several empirical challenges: physicians' decisions to batch are endogenous, studying test sequencing requires granular timestamp data rarely

available in claims databases, and establishing quasi-randomization requires detailed understanding of institutional assignment mechanisms. Our study addresses these challenges using 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). Our primary analysis leverages Mayo Clinic's rotational patient assignment system, which assigns patients to physicians via a round-robin algorithm independent of patient characteristics or physician workload (Traub et al. 2016b, Traub et al. (2016a)). We validate our findings using data from MGH, which employs a wholly different, non-random assignment mechanism. Our empirical strategy follows the "judges design" literature (Dahl et al. 2014, Dobbie et al. (2018)), exploiting variation in physicians' tendency to batch as an instrumental variable for the batching decision; full details are provided in Section 3.

1.1. Main Findings and Contributions

Whether batch ordering improves or hinders ED efficiency is an active debate among emergency physicians. Proponents argue that ordering multiple tests upfront—thereby initiating scheduling and preparation for multiple imaging workflows concurrently—reduces total processing time when multiple tests are ultimately needed; skeptics contend that this approach foregoes the information value of sequential results and leads to unnecessary testing. Our results provide causal evidence that, at least for discretionary decisions at the margin of physician judgment, batch ordering negatively impacts operational metrics. In particular, our results show that the marginal batched patient experiences a 65% increase in total ED LOS and a 69% increase in time to disposition compared to patients managed through standard practice. Furthermore, discretionary batch ordering leads to 1.2 additional imaging tests per patient encounter (an 88% increase) and a 40 percentage point increase in admission probability. These effects remain robust even after adjusting for patient and physician characteristics, laboratory testing intensity, and ED capacity. Through a decomposition analysis combining our causal estimates with controlled associations, we find that increased testing volume accounts for approximately one-quarter of the higher LOS caused by discretionary batch ordering. The substantial residual effect is consistent with increases in coordination costs and decision complexity arising from concurrent diagnostic workflows.

Examining the drivers of batch ordering, we find that physicians are more likely to batch order tests earlier in their shifts and for patients with higher acuity and more complex chief complaints. This batching tendency, however, varies systematically across physicians and persists even after controlling for patient mix, ED conditions, and observable physician characteristics. Finally, we observe that batching rates decline modestly during periods of major overcapacity (12.5% vs. 14.8%

Table 1 Descriptive Statistics of Emergency Department Encounters

Variable Patients encounters[a]	Mayo Clinic (Median [IQR]) n = 48,854	MGH (Median [IQR]) 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]
Time to disposition (min)	185 [126, 257]	—
Treatment Time (min)	160 [105, 227]	—
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 Median (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.

during normal operations), suggesting physicians may become more selective in their batching decisions under resource constraints.

Our results indicate that, despite the perceived workflow advantages of initiating multiple diagnostic workflows concurrently, batch ordering results in significantly longer processing times and increased resource utilization without corresponding improvements in patient outcomes. We validate these findings through heterogeneity analyses across seven high-batching chief complaint categories, finding consistent patterns—increased imaging without efficiency gains or quality improvements—across all clinical scenarios. These findings have important implications, highlighting that ED managers should consider strategies to reduce discretionary batching to improve diagnostic testing workflows. We shed light on these strategies and provide actionable insights on how ED managers can achieve them.

2. Related Literature

In EDs, physicians face a fundamental choice in how they sequence diagnostic imaging tests: they can order multiple tests simultaneously ("batch ordering") or order them sequentially as information accumulates. 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 standard practice, 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 heavy workload, both of which substantially influence their decision making. Workload management and carefully balancing the tradeoffs between speed and quality of care are of high importance for physicians making decisions in complex environments, such as the ED (Saghafian et al. 2018, Leppink and Hanham (2019)), and these decisions lead to multiple documented effects on physician performance. For instance, physicians may resort to ordering additional diagnostic tests when faced with limited time to interact with patients, a practice that requires less immediate critical thinking than direct clinical assessment (Batt and Terwiesch 2016, Pines (2009)).

Moreover, increased stress and frequent interruptions can disrupt systematic decision making and impair complex diagnostic reasoning (Chisholm et al. 2000, Bendoly (2011)). Frequent task switching to manage simultaneous demands can exacerbate cognitive strain and decision fatigue, prompting physicians to batch diagnostic tests and defer complex diagnostic decisions until comprehensive results are available (KC 2013, Skaugset et al. (2016), Berry Jaeker and Tucker (2020)). By ordering multiple tests upfront, physicians can reduce the cognitive strain of repeated task-switching and decision-making under uncertainty. Accordingly, a cognitive-load theory of batching suggests that the practice is more prevalent during periods of high workload or complexity.

Significant variation in ED testing and admitting practices has also been documented (Hodgson et al. 2018, Coussens and Ly (2024), Smulowitz et al. (2021)). This variation extends to batch ordering, where physicians differ systematically in their propensity to order multiple imaging tests simultaneously (Jameson et al. 2024). However, understanding the drivers and operational impact of this variation remains a significant gap in the literature. Our study aims to fill this gap by examining the factors that drive batch ordering behavior and exploiting 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 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—such as capacity limitations, varying processing times across modalities, and the inability to run different imaging types on a patient 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 also how such discretion 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 often reduce setup times and improve throughput in manufacturing settings (Fowler and Mönch 2022), applying these principles to healthcare operations introduces unique complexities. Though batching may streamline the diagnostic process by initiating multiple diagnostic workflows concurrently (reducing inter-test administrative delays even though the actual image acquisitions must occur sequentially) (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 inform the need for subsequent ones—creates a fundamental tension between operational and diagnostic efficiency that

has not been well studied. Our analysis provides novel evidence of this tradeoff, showing how different test ordering strategies affect operational metrics and clinical decision making. Our study advances this literature by providing 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.

2.3. Hypothesis Development

Building on the literature reviewed above, we develop formal hypotheses about how batch ordering affects ED operations. Though prior work has identified the mechanisms driving batching behavior and its potential consequences, the net effects remain theoretically ambiguous. Our theoretical framework centers on the fundamental tradeoff between the perceived efficiency of parallel processing and the information value of sequential testing.

2.3.1. Information Value and Test Volume The decision to batch or sequence tests fundamentally involves whether to preserve the option value of information. Sequential testing allows each test result to inform subsequent decisions, potentially eliminating unnecessary tests. When physicians batch tests upfront, they commit to a diagnostic pathway before information unfolds, forfeiting this option value.

Physicians under high workloads face cognitive strain from task switching and may batch tests to defer complex diagnostic reasoning (KC 2013, Skaugset et al. (2016)). However, this cognitive convenience comes at a cost. Without the filtering mechanism of sequential information revelation, physicians must rely solely on their initial assessment. Lam et al. (2020) identify this as a key driver of overtesting—when facing diagnostic uncertainty, physicians order comprehensive test batteries rather than allowing initial results to guide subsequent testing. Given the documented variation in physician testing intensity (Hodgson et al. 2018), with some physicians ordering twice as many tests as their peers, batching likely amplifies these tendencies by removing the natural stopping points that sequential results provide. Therefore:

Hypothesis 1. *Batch ordering will increase the total number of imaging tests performed.*

2.3.2. Processing Time and Operational Flow While batching strategies reduce setup times in manufacturing (Fowler and Mönch 2022), the ED imaging context presents unique operational constraints as noted in our review. Different imaging modalities require separate equipment and cannot be performed simultaneously (Jessome 2020). This creates a fundamental bottleneck where batched orders must still be executed sequentially, but now with a larger committed workload that cannot be adjusted based on emerging information.

Moreover, the cognitive load literature suggests that processing multiple test results simultaneously increases decision complexity (KC 2013). When physicians receive multiple results at once rather than sequentially, they must integrate more information simultaneously, potentially lengthening the diagnostic reasoning process. This "information overload" effect, combined with the additional tests ordered as predicted in H1, suggests that batching may paradoxically increase rather than decrease processing times:

Hypothesis 2. *Batch ordering will increase patient length of stay and time to disposition, as the operational constraints of imaging and increased test volume outweigh the potential benefits of parallel processing.*

2.3.3. Clinical Decision-Making and Disposition The medical literature recognizes "diagnostic momentum"—where abnormal findings, even if clinically insignificant, drive further workup and more conservative clinical decisions (Coen et al. 2022, Featherston et al. (2020)). When physicians batch order and receive multiple results simultaneously, they encounter more opportunities for incidental findings that may influence disposition decisions (Lumbreras et al. 2010, Berlin (2011)). As our review noted, physicians facing uncertainty and potential legal consequences may opt for more conservative disposition decisions (Rao and Levin 2012, Lam et al. (2020)). The simultaneous arrival of multiple test results, particularly with incidental findings, may trigger defensive medicine behaviors:

Hypothesis 3. *Batch ordering will increase hospital admission rates.*

2.3.4. Contextual Moderators The literature on physician behavior under capacity constraints consistently shows that resource scarcity forces more selective decision making (Kuntz et al. 2014, KC and Terwiesch (2009)). When EDs face severe overcrowding, the operational pressures documented in our review intensify. Under these conditions, physicians may reserve batching for cases where it is clinically essential rather than convenient:

Hypothesis 4. *The effects of batch ordering on LOS and test volume will be attenuated under conditions of major ED overcapacity.*

These hypotheses provide testable predictions that we examine using our quasi-experimental design. By leveraging variation in physician batching tendency under random patient assignment, we identify whether these theoretical mechanisms manifest in actual ED operations.

3. Setting, Data, and Models

3.1. Empirical Setting

Our study uses data from two large U.S. EDs: the Mayo Clinic of Arizona and 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. Our primary analysis, however, focuses on the Mayo Clinic data due to its unique random patient-physician assignment, which enables stronger causal inference. Specifically, the Mayo Clinic ED employs a computerized rotational patient assignment algorithm that addresses many empirical challenges in healthcare settings; see, e.g., Traub et al. (2016a), Traub et al. (2016b), Traub et al. (2018). The system assigns patients on a completely rotational basis to physicians 60 seconds after registration in the electronic health record system. At shift start, each physician receives four consecutive patients to establish an initial load, after which they enter rotation with other on-duty physicians. The rotation order is predetermined by the ED scheduler and varies across shifts to ensure fairness over time. Critically, these assignments are based solely on arrival time—the algorithm does not consider patient demographics, chief complaint, Emergency Severity Index score, physician workload, or the acuity of recently assigned patients. To maintain system integrity, physicians receive no new patients during their final 120 minutes and are capped at 18 patients per shift.

Our empirical tests show that this rotational mechanism achieves the quasi-randomization necessary for causal inference. Unlike settings where patient-physician matching may be influenced by triage decisions, physician preferences, or informal routing practices, the Mayo Clinic's algorithmic assignment removes discretion from the matching process. We establish that, conditional on arrival time, patient-physician matching is effectively random—a critical requirement for our identification strategy that distinguishes our study from observational analyses where endogenous matching could confound the effects of physician discretion.

The data include information on the timing of test orders, test results, patient disposition, and other important triage metrics and demographic features. We focus on imaging tests (x-rays, contrast CT scans, non-contrast CT scans, and ultrasound) because, unlike laboratory tests, these tests cannot be run simultaneously on a given patient due to the different equipment and settings required. Therefore, the operational implications of batch ordering imaging tests are more pronounced. We exclude MRI from our Mayo analysis because an institutional policy requires either inpatient admission for urgent MRIs or outpatient ordering for non-urgent cases, resulting in negligible ED MRI use. MGH does not have this policy, so we included MRI in its generalizability analysis.

3.2. Data

Our primary data comes from the Mayo Clinic of Arizona ED, a tertiary care hospital without obstetrical services, an inpatient pediatrics unit, or a trauma designation. During the study period (October 6, 2018, through December 31, 2019), the ED recorded 48,854 visits, managed across 26 treatment rooms and up to 9 hallway spaces. The department is staffed exclusively by board-eligible

or board-certified emergency physicians (EPs), a rare yet ideal setting for our study. Many EDs are staffed by a mix of EPs and non-EPs (e.g., Nurse Practitioners and Physician Assistants), both of whom are responsible for ordering tests, which may introduce confounding factors. The Mayo Clinic ED is unique in that only EPs can order tests, eliminating the potential for confounding by provider type. Furthermore, as mentioned earlier, the ED uses a randomized patient-to-EP assignment, eliminating some other potential sources of confounding.

We conducted a retrospective review of the comprehensive ED operations 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 Figure EC.1.1. LOS is measured from arrival to departure from the ED. During periods when inpatient beds are filled (i.e., when patients requiring inpatient admission must "board" in the ED due to lack of inpatient beds), the Mayo Clinic converts 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" boarding bed within the ED [1].

To improve power in our analyses, we drop encounters with rare reasons for visit (defined as those with fewer than 1,000 total encounters) and complaints for which a batch order occurs less than 5% of the time across all patients or for which no imaging is ordered. Since batch orders are rare in these cases, our physician batch tendency instrument could suffer from a weak-instrument problem if we included them. Examples of complaints dropped include skin and urinary complaints, as well as other complaints in which multiple imaging modalities are unlikely to be indicated. Figure EC.3.1 provides a CONSORT-style flow diagram detailing each exclusion step and the corresponding number of observations removed, and Table EC.3.1 compares characteristics of excluded versus included encounters.

Excluding these conditions does not introduce selection bias: Mayo Clinic's random assignment ensures physicians see the full spectrum of acuity and complaints regardless of our sample restrictions. We focus on encounters where the batching decision is both consequential and discretionary—precisely the population where our LATE provides actionable policy guidance. Though our analytical sample represents 24% of total ED encounters, these cases account for 41% of imaging resource utilization, underscoring their operational significance. Finally, to estimate a precise measure of physician-level batch tendency, we restrict our sample to encounters involving only full-time

[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.

physicians. Our final sample includes 11,679 encounters, with chief complaints from the following categories: Neurological Issue, Abdominal Complaints, Fevers/Sweats/Chills, Falls/Motor Vehicle Crashes/Assaults/Trauma, Dizziness/Lightheadedness/Syncope, Extremity Complaints, and Fatigue/Weakness.

3.2.1. Treatment Variable Our treatment variable, $Batched_{i,t}$, is an indicator that equals 1 if patient i had tests batch ordered during an ED encounter 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. Again, this contrasts with standard practice, where a single test is ordered, and subsequent tests are ordered in sequence as needed.

We define "batching" in line with standard emergency medicine practices and focus on batches that include two or more different imaging modalities ordered within a 5-minute window from the timestamp of the first imaging order (Su et al. 2025, Jameson et al. (2024)). We focus on early batching (within 5 minutes) because this represents the moment of maximum diagnostic uncertainty when physicians must decide their testing strategy before clinical information unfolds. Physicians cannot know *ex ante* which patients will ultimately require multiple tests, making early batching a discretionary choice based on practice style rather than clinical necessity. 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 where the physician orders different imaging tests because such tests cannot be done in a single scanning session (due to differences in equipment and setting). Encounters where a single test precedes subsequent batched tests (1.91% of multi-test cases) are classified as standard care in our primary analysis, as the physician has initiated sequential information gathering before placing additional orders. Sensitivity analyses conducted around this time window, batch size threshold, and the timing of the batch show that our results are robust to variations in these values.

3.2.2. Dependent Variables The primary outcomes of interest are the ED's efficiency and effectiveness measures. We measure patient LOS in two 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 excludes delays related to inpatient bed availability, providing a more transparent measure of ED operational efficiency. Second, we measure a patient's total time in the ED from arrival until physical departure (Lim et al. 2024). This total time for admitted patients includes the duration until transfer to an "inpatient" bed, whether in the main hospital or designated ED areas converted for inpatient use, encompassing boarding time and discharge processing (Feizi et al. 2023). Given the documented right-skewness of ED time metrics (Song et al. 2015), we

log-transform both time measurements to approximate normality (Brown et al. 2005, Saghafian et al. (2024)), meeting the assumptions required for our regression analyses. As a robustness check, we also examine treatment time (from physician contact to disposition), excluding both waiting room delays and post-disposition boarding; results are qualitatively unchanged and are reported in EC.10.

Beyond time-based metrics, we examine resource utilization through the number of distinct imaging tests performed during each ED encounter. This count variable captures tests actually completed with documented results, not merely ordered—an important distinction since ordered tests may occasionally be cancelled before completion[2]. This measure helps us understand how batch ordering practices influence diagnostic workload.

To assess care quality, we track whether patients return to the ED within 72 hours of discharge from their initial visit and require hospital admission (Lerman and Kobernick 1987). We focus on returns requiring admission rather than any 72-hour return because this measure better captures actual quality failures. Some ED revisits are planned or expected—patients may be instructed to return for wound checks, suture removal, or if symptoms persist after initial treatment. Returns requiring admission, however, are more likely to signal potential issues with initial treatment decisions, premature discharges, or missed diagnoses. For patients admitted during their index visit, the 72-hour window begins at hospital discharge rather than ED departure, ensuring fair comparison across disposition types. We verify robustness to the broader any-return measure in EC.10.

These measures allow us to evaluate ED performance across three critical dimensions: operational efficiency through time-based measurements, resource utilization via imaging tests performed, and care quality through return visit patterns. By examining these outcomes together, we can assess how batching behaviors affect the efficiency and effectiveness of care delivery.

3.3. Identification Strategy

Our goal is to estimate the causal effect of discretionary batch ordering diagnostic imaging tests on patient outcomes (formalized in Eq. (4) below). A key challenge is that the decision to batch is endogenous: physicians may be more likely to order multiple tests simultaneously for patients with greater unobserved severity, which independently affects outcomes such as length of stay and return admissions. To address this endogeneity, we exploit the quasi-random assignment of patients to ED physicians who differ in their underlying tendency to batch order tests. This "batch tendency" serves as an instrumental variable, leveraging the rotational assignment mechanism in the ED to

[2] Test cancellations after ordering are extremely rare and face substantial operational barriers. Once the radiology department acknowledges an order, cancellation requires physicians to call and request that the department "push back" the imaging order. Furthermore, radiology departments often coordinate between modalities (e.g., CT and ultrasound) so patients move directly from one scanner to another.

generate exogenous variation in batching. We first describe the construction and validation of this instrument, then present our IV specification and identifying assumptions.

To construct this instrument, our empirical strategy closely follows the literature that relies on the quasi-random assignment of agents to cases, often referred to as the "judges design." Papers in this literature typically exploit variation in judges' sentencing leniency within the same court. Similarly, we explore variation in batching across physicians in the same ED using a measure we term "batch tendency." We use each physician's residualized leave-out average batch rate to measure physician batch tendency. We use this residualized measure of physician batch tendency because, if certain physicians are more likely to work afternoon or weekend shifts (as Figure EC.1.1 shows are the busiest shifts), the simple leave-out mean will be biased. A residualized measure of physician batch tendency accounts for this potential selection. This measure is derived from two steps following a similar approach used in other applications (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 on date t. Fixed effects include year-month fixed effects, α_{ym}, to control for time- and season-specific variation in batching, hospital-specific policies (e.g., initiatives to eliminate excess testing during a flu season), and 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}[3]. 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. Our primary specification adds further precision controls, including laboratory tests ordered, physician characteristics, and ED capacity, as detailed in Section 3.4. As stated in Section 3.1, these controls are more than 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:

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

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

where $\hat{\varepsilon}_{i'} = Batch_{i'} - \hat{Batch}_{i'}$ is the residual from Eq. (1); $\mathbb{J}$ is the set of all ED encounters treated by physician j; and $N_{-i,j} = |\{\mathbb{J} \backslash i\}|$, the number of cases that physician has seen that year, excluding patient i. This leave-out mean eliminates the mechanical bias that arises when patient i's case enters 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, and time of day.

Figure 2 provides empirical verification that, while the decision to batch depends on patient characteristics, our measure—batch tendency—is plausibly exogenous. The left panel uses a linear probability model to test whether encounter, patient, ED, and physician characteristics predict the batching decision, controlling for shift-level fixed effects. As expected, patient characteristics strongly predict batching decisions; for instance, patients with Falls/Assaults/Trauma complaints are 16.2 percentage points more likely to be batched compared to similar patients under similar ED capacity. The right panel assesses whether these same characteristics predict assignment to physicians with different batch tendencies. Importantly, we find that patient characteristics do not significantly predict assignment to high- or low-batch-tendency physicians. The coefficients are near zero, with confidence intervals crossing zero for all patient characteristics, confirming that, conditional on shift-fixed effects (which account for mechanical rotation), the assignment of patients to physicians with different batching tendencies is effectively random. This validates the rotational assignment mechanism and establishes batch tendency as an exogenous source of variation for identifying causal effects.

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 shown, the probability of batching increases approximately linearly and monotonically with our tendency measure (see Table 2 for more formal results).

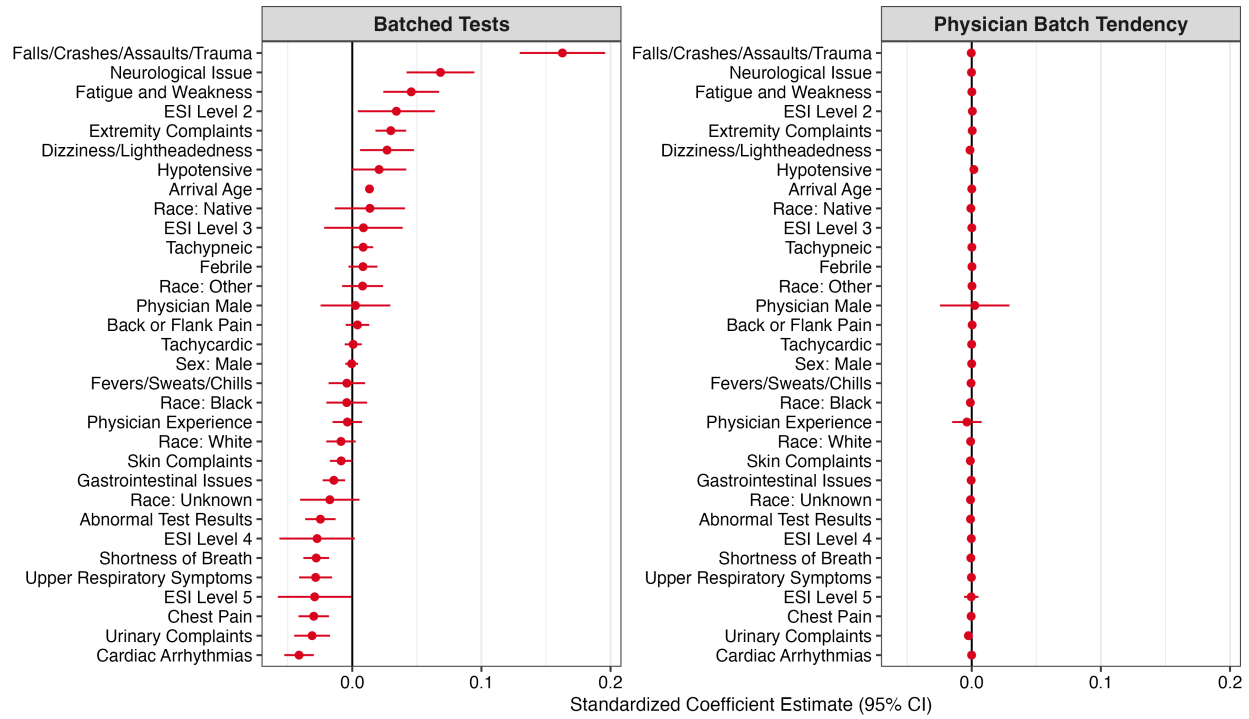
Together, we observe that batch tendency likely meets the criteria for a valid IV. In the next section, we formalize our IV analysis and establish its validity.

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 batching on outcomes, Y_i, we estimate the following 2SLS equations:

Figure 2 Batch Tendency by Patient Characteristics



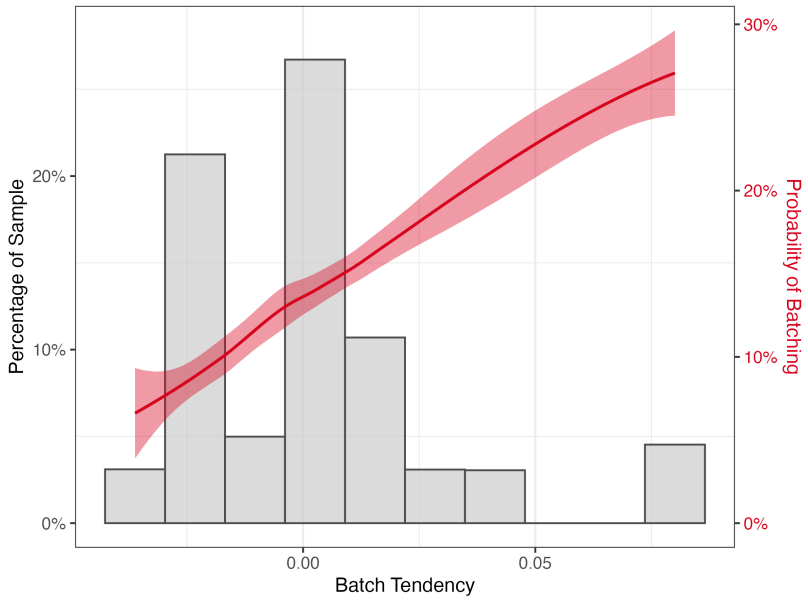
Notes: This figure tests quasi-random assignment of patients to physicians in the Mayo Clinic ED. The left panel shows how patient characteristics predict batching decisions. The right panel shows that these same characteristics do not predict assignment to physicians with different batch tendencies.

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

In all cases, Y_i represents our primary 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 Table EC.2.1). Note that the subscript i indexes patient encounters, with the temporal dimension captured through fixed effects in $\mathbf{X_i}$. The variable $Batched_i$ may be endogenous; for example, injury severity may be unobserved and correlated with the need to run multiple tests, length of stay, and the likelihood of returning with admission. Hence, we instrument $Batched_i$ with the assigned physician j's underlying tendency to batch, $BatchTendency_{i,j}^{phys}$. Standard errors are heteroskedasticity-robust (EC.13 provides additional details).

Table 2 presents first-stage results from Eq. (4). Column 1 of Table 2 presents the overall mean and standard deviation of the dependent variable (Batched) in the baseline sample. Column 2 reports results only with year-month and day-of-week-time-of-day fixed effects. Column 3 adds

Figure 3 Distribution and First Stage of Instrument



Notes: 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 on the right y-axis. Ninety-five percent confidence bands are also shown.

our baseline patient and hospital condition controls. Consistent with Figure 3, 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 patients to physicians with different batching tendencies.

Furthermore, the batch tendency measure reasonably predicts the batching decision, and the IV is not weak (F-stat = 171.9). For example, across all controls (column 3), our results show that a patient assigned to a physician at the 90th percentile of batch tendency (0.023) relative to a physician at the 10th percentile (-0.024) is 8.2 percentage points more likely to have their tests batched. The coefficient is greater than one because the BatchTendency instrument has a narrow range centered near zero (10th to 90th percentile spans only 0.047 units). A one-unit change in the instrument is therefore far outside the observed range, making per-unit interpretation uninformative. Instead, we interpret the effect over realistic variation: moving from the 10th to 90th percentile of batch tendency increases batching probability by 8.2 percentage points, as shown above.

3.3.2. UJIVE Construction Leniency designs like ours, where many providers generate the identifying variation, can suffer from many-instrument bias in conventional 2SLS: the constructed

Table 2 First-Stage Results: Batch Tendency and Batching

	Sample Mean	Batched	
	(1)	(2)	(3)
Batch Tendency	0.137	1.837***	1.752***
	(0.343)	(0.138)	(0.134)
Controls			
Necessary controls	—	Yes	Yes
Precision controls	—	No	Yes
Adj. R^2	—	0.022	0.74
F-stat	—	177.5	171.9
Observations	11,679	11,679	11,679

Notes: This table reports first-stage results for the regression of batch tendency on the likelihood of batching. Column 1 reports the overall mean and standard deviation of the dependent variable (Batched) in the baseline sample. Column 2 includes quasi-random assignment necessary controls for year-month and day of week-time of day. Column 3 adds baseline precision controls, including patient characteristics (age, ESI-complaint, race, sex, and vital signs such as tachycardia, tachypnea, febrile status, and hypotensive status), laboratory tests ordered, ED capacity level, and physician characteristics (physician sex, experience, and hours into shift). Robust standard errors are heteroskedasticity-robust. $^{***}p < 0.01$.

first-stage fitted values mechanically reuse each observation's own treatment status, creating a small-sample correlation between the instrument and the structural error (Kolesár et al. 2015, Goldsmith-Pinkham et al. 2025). This bias can be substantial even when first-stage F-statistics appear adequate.

To address this concern, we implement the unbiased jackknife instrumental variables estimator (UJIVE) as a robustness check for our main 2SLS specification. While our primary design uses a single continuous instrument—physician batch tendency—UJIVE uses the full set of physician indicators, Z_j, as instruments. For each observation i treated by physician j, UJIVE constructs a leave-one-out predicted instrument:

$$\hat{\ell}_{i,-i} = Z_j' \hat{\pi}_{-i} \tag{5}$$

where $\hat{\pi}_{-i}$ is estimated from the first-stage regression of *Batched* on physician indicators using all observations except i. Because $\hat{\pi}_{-i}$ excludes observation i's own treatment status, the predicted instrument $\hat{\ell}_{i,-i}$ is orthogonal to the structural error by construction. This eliminates the mechanical correlation that contaminates conventional many-instrument 2SLS while preserving consistency for the local average treatment effect.

This estimator is asymptotically unbiased in many-instrument leniency designs and yields valid standard errors even when conventional 2SLS understates uncertainty. If our 2SLS results were

driven by many-instrument bias, we expect UJIVE estimates to diverge substantially. As we show in Section 4, UJIVE estimates closely track our 2SLS results across all outcomes, increasing confidence that many-instrument bias is not driving our findings.

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

Our instrumental variable approach, which aims to recover the causal effect of batch ordering diagnostic tests, 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 the instrument to influence the outcome of interest only through its effect on test batching. This assumption is untestable, but we take it seriously and address it through multiple strategies. We expand our precision control set to include laboratory tests ordered, physician characteristics (experience, gender, hours into shift), and ED capacity levels in addition to patient complaint, sex, race, acuity, and vital signs. Additionally, we test this assumption by performing a placebo check for rarely batched complaints (EC.8) and various robustness checks in Section 4.8, including sensitivity analyses (EC.9) that allow for plausible violations of the exclusion restriction (Conley et al. 2012).

To the extent that any residual exclusion restriction concerns remain, our reduced-form estimates can be interpreted as the causal effect of being assigned to a high-versus low-batch-tendency physician—a policy-relevant parameter for ED managers considering interventions around physician feedback or training.

Finally, the monotonicity assumption is necessary for interpreting the IV coefficient estimates as LATEs when there are heterogeneous treatment effects. It requires that any patient batched by a low-batch-tendency physician also be batched by a high-batch-tendency physician. The literature that leverages the judges' design typically performs two informal tests to assess monotonicity. The first provides that the first stage should be weakly positive for all subsamples (Dobbie et al. 2018). The second asserts that the instrument constructed by omitting a particular subsample has predictive power for that same subsample (Bhuller et al. 2020). Figure EC.7.1 presents both of these tests for various subsamples of interest. In the left panel, our residualized measure of batch tendency is consistently positive and sizable across all subsamples, consistent 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.

3.3.4. Details on Our LATE Estimates Our 2SLS estimates represent the LATE of batch ordering for "compliers"—patients whose testing strategy depends on the assigned physician's practice style. This effect compares early batching to standard practice, which includes both sequential ordering and single tests. Though this involves a composite counterfactual, it provides the policy-relevant parameter: the effect of encouraging comprehensive upfront testing versus allowing diagnostic information to guide testing decisions for patients at the margin of clinical discretion—those whose testing strategy is not dictated by apparent clinical necessity but rather depends on physician practice style and judgment. 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. (2014).

Specifically, compliers are defined as patients whose batched status 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 assigned physician. Following Dahl et al. (2014) and Dobbie et al. (2018), we define the highest batch tendency physicians as those at the 99th percentile of batch tendency and the lowest as those at the 1st percentile. Using a local linear regression of batching on batch tendency, we predict the probabilities of being batched under the most aggressive ($\bar{z}$) and most conservative ($\underline{z}$) physicians.

Approximately 20 percent of patients in our sample are "compliers," meaning they would have received batched tests if assigned to a high-batch-tendency physician but received standard practice otherwise. In comparison, 7 percent of patients are "always-takers," meaning they would receive batched tests regardless of the physician's batch tendency, and 73 percent are "never-takers," meaning they would never receive batched tests regardless of the physician's batch tendency. Complier characteristics reveal that marginal patients—those at the boundary of physician discretion—tend to present with normal vital signs, standard laboratory workups, and are disproportionately likely to have trauma or neurological complaints. This pattern is consistent with clinical intuition: these presentations involve diagnostic uncertainty where physician practice style most influences testing decisions. EC.4 provides additional details on complier estimation and characteristics.

4. Results and Discussion

4.1. Reduced-Form Results

In this sub-section, we explore the causal influence of physician batch tendency on patient outcomes and resource utilization in the ED. We find statistically and operationally significant effects

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	1.140^{***}	1.046^{***}	2.173^{***}	-0.0257
	(0.1592)	(0.1350)	(0.2103)	(0.0341)
Scaled coeffecient: 10th $\rightarrow$ 90th pct. ($\Delta = 0.050$)	0.0576	0.0525	0.1089	-0.00129
Mean dependent variable	5.248	5.505	1.450	0.0121
	(0.493)	(0.449)	(0.623)	(0.109)
Necessary controls	Yes	Yes	Yes	Yes
Precision controls	Yes	Yes	Yes	Yes
Adj. R^2	0.2109	0.2834	0.1346	0.0103
Observations	11,679	11,679	11,679	11,679

Notes: This table reports reduced-form estimates of the relationship between physician batch tendency and patient outcomes. The second row rescales coefficients to represent the effect of moving from the 10th to the 90th percentile of batch tendency ($\Delta = 0.050$). $^{***}p < 0.001$. Coefficients on log-transformed dependent variables can be interpreted as percentage changes using the formula $(\exp(\beta) - 1) \times 100\%$.

of assignment to a high-batch-tendency physician on every outcome except 72-hour return with admission (Table 3). Scaling our coefficients by the difference in tendency going from the lowest decile to the highest decile in physician tendency—equal to a 5.0 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 a 5.8% increase in time to disposition, a 5.3% increase in LOS, and an additional 10.9 imaging tests ordered per 100 patient encounters. These findings highlight ED physicians' substantial role in putting patients on a path toward longer LOS and increased resource utilization[4]

The fact that our primary outcomes respond strongly to physician batch tendency suggests that batching is 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. EC.8 attempts to distinguish between and identify the mechanisms behind the observed reduced-form effects. This mediation analysis provides a crucial step toward a well-identified IV analysis, which we present in Section 4.2 using both 2SLS and UJIVE estimators.

[4] For reduced-form results for a less strict sample see, e.g., Jameson et al. (2024), Hodgson et al. (2018), and the references therein.

4.2. Instrumental Variables Estimation

Next, we examine the effects of batch ordering imaging tests using the 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 2SLS and UJIVE estimates of the impact of batching on key operational metrics. Column 1 reports the dependent variable means for patients managed through standard practice. Columns 2 and 3 report 2SLS estimates using our physician batch tendency instrument, without and with precision controls, respectively. Columns 4 and 5 report UJIVE estimates, which address potential many-instrument bias by using leave-one-out predictions from the complete set of physician indicators.

The UJIVE estimates with full controls (Column 5)—our preferred specification—indicate that discretionary batching substantially increases ED length of stay. The marginal batched patient experiences a 65% increase in total ED length of stay and a 69% increase in time to disposition compared to patients managed through standard practice. Notably, the estimates attenuate when precision controls are added (comparing Columns 2-3 and Columns 4-5), confirming that our original specification captured some correlation with general diagnostic intensity. However, large and significant effects persist.

Discretionary batching also leads to more intensive diagnostic testing. The marginal batched patient receives 1.2 additional imaging tests (completed studies with documented results), representing an 88% increase from the mean for patients managed through standard practice. This increased testing intensity does not appear to improve care quality: effects on 72-hour returns with admission are small, negative, and statistically indistinguishable from zero. These findings suggest that discretionary batching results in additional delay-inducing tests that do not add diagnostic value. Put differently, standard practice offers an important benefit: the information obtained from initial tests reduces the need for 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 estimates reveal that batching leads to significant increases in X-ray utilization: with full controls, batched patients are approximately 96 percentage points more likely to receive an X-ray ($p < 0.001$). Effects on other modalities—ultrasound, CT without contrast, and CT with contrast—are smaller and not statistically significant after adding precision controls. This pattern suggests that physicians who batch construct comprehensive workups primarily by adding quick, low-cost imaging (X-rays) rather than selectively ordering expensive advanced imaging.

Panel C of Table 4 presents estimates of batching's impact on disposition decisions. The UJIVE results with full controls indicate that discretionary batching increases the admission probability

Table 4 Effect of Batching Tests on Patient Outcomes

	Sequenced mean	2SLS		UJIVE	
	(1)	(2)	(3)	(4)	(5)
Panel A. Primary Outcomes					
Log time to disposition	5.237	0.659***	0.651***	0.583***	0.522***
	(0.499)	(0.103)	(0.101)	(0.189)	(0.177)
Log LOS	5.490	0.717***	0.597***	0.653***	0.503***
	(0.456)	(0.094)	(0.088)	(0.158)	(0.144)
Number of distinct imaging tests	1.335	1.385***	1.241***	1.316***	1.174***
	(0.572)	(0.118)	(0.116)	(0.126)	(0.119)
72-hour return with admission	0.012	-0.0137	-0.0146	-0.0079	-0.0039
	(0.110)	(0.018)	(0.019)	(0.020)	(0.022)
72hr return	0.030	-0.0512	-0.0536	-0.0440	-0.0396
	(0.170)	(0.029)	(0.031)	(0.032)	(0.034)
Panel B. Test Types					
X-ray	0.576	0.943***	0.989***	0.960***	0.959***
	(0.494)	(0.100)	(0.101)	(0.116)	(0.117)
Ultrasound	0.171	0.160**	0.087	0.164*	0.087
	(0.377)	(0.076)	(0.073)	(0.082)	(0.078)
CT without contrast	0.400	0.102	0.062	0.052	0.053
	(0.490)	(0.095)	(0.086)	(0.112)	(0.102)
CT with contrast	0.187	0.180*	0.102	0.140	0.075
	(0.390)	(0.078)	(0.076)	(0.087)	(0.079)
Panel C. Disposition					
Admission	0.279	0.419***	0.404***	0.424***	0.398***
	(0.449)	(0.096)	(0.088)	(0.103)	(0.090)
Necessary controls	—	Yes	Yes	Yes	Yes
Precision controls	—	No	Yes	No	Yes
Observations	11,679	11,679	11,679	11,679	11,679

Notes: Column 1 reports means for standard care patients (standard deviations in parentheses). Columns 2–3 report 2SLS results using physician batch tendency as the instrument. Columns 4–5 report UJIVE estimates using ED provider identifiers as many weak instruments. All models include day-of-week and time-of-day fixed effects, as well as month-year fixed effects. Precision controls are described in the text. Standard errors are heteroskedasticity-robust. Coefficients on log-transformed dependent variables can be interpreted as percentage changes using the formula $(\exp(\beta) - 1) \times 100\%$

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

by 40 percentage points. This also means that the impact of batching in the EDs likely spills over to other parts of hospitals (e.g., inpatient units), increasing their patient volumes. Increased

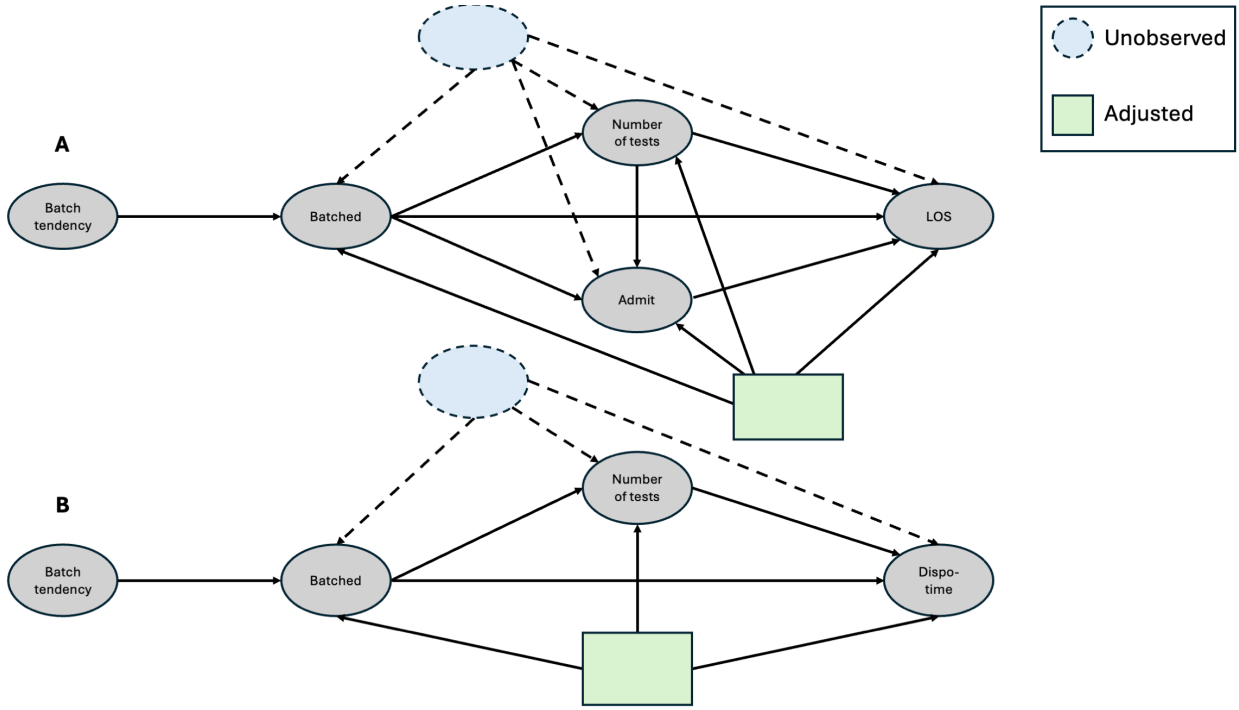


Figure 4 Directed Acyclic Graph (DAG) of Hypothesized Mechanisms

patient volume, in turn, is known to induce behaviors such as speed-up, which can harm other quality-of-care metrics, such as 30-day mortality (Song and Saghafian 2019).

These results paint a consistent picture of the operational implications of batching. Discretionary batching results in more comprehensive diagnostic workups but also significantly longer processing times and a higher admission probability, without measurable improvements in short-term quality outcomes. The stability of estimates across 2SLS and UJIVE specifications, and the expected minor attenuation when adding precision controls, strengthens confidence that these findings reflect the causal effect of imaging-specific batching behavior rather than general physician diagnostic style.

4.3. Potential Mechanism: Decomposition Analysis

To understand the channels through which batching increases LOS and time to disposition, we decompose the total causal effect into components attributable to imaging volume and admission decisions. These variables are natural candidates: our IV estimates establish that batching causally increases both (Table 4), and each has clear theoretical reasons to extend processing times. Figure 4 formalizes these pathways, with admission relevant only for LOS (Part A) since disposition time by definition excludes post-disposition delays (Part B).[5]

We employ two complementary approaches. Our first multiplies the IV-estimated causal effect of batching on each mediator ($\hat{\alpha}_k$) by the controlled OLS association between that mediator and the

[5] Of note, a change in admission decision (discharge vs. admit) in the ED can influence a patient's LOS (Part A) but not their time to disposition (Part B), since the latter by definition excludes delays after the disposition decision.

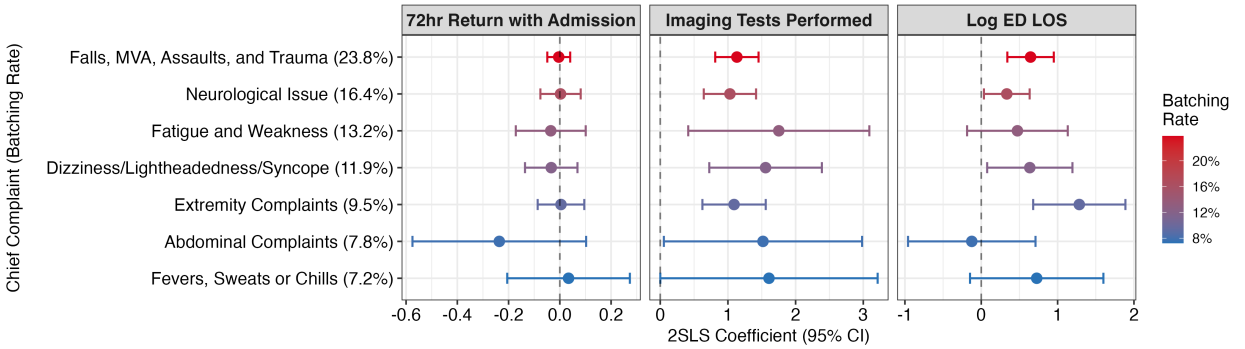
outcome ($\hat{\beta}_k$). The implied indirect effect through channel k is $\hat{\alpha}_k \times \hat{\beta}_k$, and the residual is the difference between the total IV effect and the sum of indirect effects. This approach requires that the OLS associations approximate the causal mediator–outcome relationships conditional on our controls—an assumption made more plausible by the breadth of our control set but not directly testable. We therefore interpret the decomposition as an approximate accounting of magnitudes rather than a precise causal partition. Our second approach examines how the reduced-form coefficient on batch tendency attenuates when mediators are added as controls. While adding post-treatment variables introduces bias in the strict causal sense, the degree of attenuation provides model-free evidence about the importance of each channel—an approach commonly used in the judge-leniency literature as complementary evidence for mechanism identification (e.g., Dobbie et al. 2018, Doyle et al. 2015).

Table 5 presents both sets of results. Panel A shows the product-of-coefficients decomposition. For LOS, approximately 26% of the total causal effect (0.154 of 0.597 log points) is consistent with operating through increased imaging volume: batching causally increases imaging by 1.24 tests (from Table ref{tab:results_table}), and each additional test is associated with a 12.4% increase in LOS conditional on controls.[6] The admission channel accounts for a smaller share (approximately 3%), likely attenuated by conditioning on complaint-by-ESI interactions that strongly predict both admission and LOS. The remaining 71% constitutes a residual capturing other channels: coordination delays from managing concurrent imaging workflows, cognitive complexity of processing simultaneous results, and scanner scheduling bottlenecks. For disposition time, where admission is not a mediator, the imaging channel accounts for 6% of the total effect, with the larger residual consistent with batching affecting decision-making speed through information complexity rather than test count alone. Importantly, this residual does not suggest an exclusion restriction violation. The residual concerns the treatment not the instrument, reflecting channels through which batching affects LOS beyond imaging count and admission, which is expected given the crudeness of these mediators relative to the full set of operational mechanisms batching engages.

Panel B presents the reduced-form attenuation results, which corroborate the decomposition. When imaging volume is added to the baseline reduced-form specification, the batch tendency coefficient on LOS falls from 1.046 to 0.786, a 24.9% reduction. Further adding admission produces a total attenuation of 26.5%—closely matching the 29% combined indirect share from Panel A. For disposition time, adding imaging attenuates the coefficient by 8.0% (from 1.140 to 1.049), consistent with the 6% share from the product-of-coefficients approach. The convergence across these two independent approaches strengthens confidence in the decomposition.

[6] These total effects are consistent with our preferred UJIVE estimates in Table 4; we use 2SLS here because the decomposition is constructed around the batch tendency instrument.

Figure 5 Heterogeneity in Batch Ordering Effects by Chief Complaint



Notes: Each panel shows 2SLS estimates of batching effects for different chief complaints, ordered by batching rate. Error bars show 95% confidence intervals. All models include full controls.

Together, both approaches provide convergent evidence that increased imaging volume is a quantitatively important—though not exclusive—channel through which batching extends processing times. These findings reinforce that the operational costs of discretionary batching arise substantially from the increased diagnostic activity it generates, consistent with the "information gain" advantage of standard practice, which uses initial test results to filter out unnecessary subsequent testing.

4.4. Heterogeneity Across Clinical Presentations

Our sample includes seven chief complaint categories that vary substantially in both clinical complexity and prevalence of batching, ranging from 7% for fevers to 24% for polytrauma cases. To assess whether the effects of batching depend on clinical presentation, we estimate our primary 2SLS specification separately for each complaint category. This analysis allows us to determine whether the operational inefficiencies we document are driven by specific clinical scenarios or represent a more general phenomenon.

Figure 5 presents these results. We frame this as an exploratory analysis given multiple-hypothesis-testing concerns across seven categories and three outcomes. Rather than emphasizing statistical significance in subgroups, we focus on whether any clinical scenario shows qualitatively different patterns in direction or magnitude.

The results reveal striking consistency across clinical presentations. Imaging effects are positive across all seven complaint categories, ranging from 1.03 to 1.76 additional tests, with six of seven showing statistically significant increases despite smaller subgroup sample sizes. Notably, Falls/MVA/Assaults/Trauma—our most complex category with the highest batching rate (24%)—shows a 1.13-test increase, comparable to simpler presentations. Time effects similarly show no evidence of efficiency gains in any category; the complaint-specific effects on log LOS closely track

the pooled estimate and are uniformly non-negative. Quality effects, measured by 72-hour returns with admission, are small and statistically indistinguishable from zero across all seven complaints.

This consistency suggests that the operational inefficiencies of discretionary batching are not artifacts of specific clinical scenarios. Even in complex presentations where comprehensive imaging might seem most justified, standard practice — preserving diagnostic flexibility yields operational performance that is similar or better without compromising care quality. These findings simplify policy recommendations: interventions to reduce unnecessary batching appear warranted across the full spectrum of ED presentations, not only for particular complaint types.

4.5. Heterogeneity Across ED Capacity Status

Given that ED capacity constraints significantly influence operational decisions, we examine whether the effects of batching vary across different capacity levels. Following the Mayo Clinic ED's internal guidelines[7] (Table EC.2.1), we categorize each encounter into three categories: occurring during normal operations, minor overcapacity, or major overcapacity. While our primary analysis controls for capacity status, stratifying by this variable allows us to assess Hypothesis 4—whether the effects of batching are attenuated under severe resource constraints.

Table 6 presents 2SLS estimates stratified by ED capacity status. We observe that batching rates decline modestly as capacity constraints intensify: from 14.8% during normal operations to 12.7% under minor overcapacity and 12.5% under major overcapacity. This pattern suggests that physicians become somewhat more selective in their batching decisions when facing resource constraints.

However, we do not find statistically significant differences in the effects of batching across capacity levels. During normal operations, batching increases LOS by approximately 83% (Column 1), while under minor overcapacity, the effect is similar at 90% (Column 2). Given major overcapacity, the point estimate is smaller (36%) and not statistically significant. However, the confidence intervals are wide due to the smaller sample size (n=1,083), and we cannot reject the hypothesis that the effect equals that observed under normal operations. Similarly, imaging volume effects remain large and significant across all capacity levels, ranging from 1.17 to 1.91 additional tests per batched patient.

These results do not support Hypothesis 4. While batching rates decline modestly under overcapacity—suggesting physicians exercise somewhat greater selectivity—the estimated effects of batching on outcomes remain statistically indistinguishable across capacity levels. The operational inefficiencies associated with discretionary batching persist regardless of ED utilization. From a policy perspective, this suggests that interventions to reduce unnecessary batching are warranted across all operational contexts, not only during periods of normal capacity.

[7] The ED of our other partner hospital (MGH) does not follow these guidelines. Hence, we did not include these analyses in our study of the effects of batching at MGH (see Section 4.7 for our analysis using MGH data).

4.6. Determinants of Image Batching

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

$$Y_{i,j} = \beta_0 + \beta_1 \mathbf{MD_j} + \gamma Capacity + \alpha \mathbf{X_i} + \epsilon_i \tag{6}$$

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 distinct imaging tests performed 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 Table EC.2.1. $\mathbf{X_i}$ is the vector of patient covariates described in the previous section and in Figure 2. Table 7 presents the results.

We find that physician experience is associated with modest reductions in both batching and overall imaging volume. Each additional year since residency is associated with a 0.1 percentage-point decrease in the probability of batching ($p < 0.05$) and 0.3 fewer imaging tests per 100 encounters ($p < 0.001$). Physician gender shows no significant relationship with either batching or test volume.

The timing within a physician's shift significantly influences batching decisions. For each additional hour into the shift, the likelihood of batching decreases by 0.4 percentage points ($p < 0.05$), though hours into the shift show no significant relationship with overall imaging volume. 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 influence test ordering patterns. Under minor overcapacity, we observe significant decreases in both batching (1.7 percentage points, $p < 0.01$) and imaging volume (2.6 fewer tests per 100 encounters, $p < 0.05$). Given major overcapacity, point estimates suggest reductions relative to normal operations, but these effects are not statistically significant—likely due to the smaller number of encounters during these periods (n=1,083). This pattern is consistent with physicians becoming more selective in their ordering behavior when facing resource constraints, though the relationship is modest in magnitude.

4.7. 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 patients are randomly assigned to physicians upon arrival through a rotational system, the MGH ED employs a different patient assignment mechanism. At MGH, patients are triaged into different care areas (e.g., urgent care, fast track, observation) based on acuity and presenting complaints, then assigned to physicians based on availability within those areas rather than through random rotation. To address this non-random assignment and potential selection bias, we adjust our instrumental variable strategy to account for these differences by including additional covariates for care area assignment, acuity level, and presenting complaints in both stages of our 2SLS and instrument construction, thereby accounting for the sorting of patients into different ED zones. While this approach cannot guarantee the same level of causal identification as making use of Mayo Clinic's randomized system, it provides a more robust comparison of the effects of batching on patient outcomes across different ED settings

After adjusting for institutional differences and using the same exclusion criteria we used with Mayo, we find strong evidence that our key findings generalize to the MGH setting. The 2SLS results in Table 7 suggest that batching leads to a 40.5% 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 mainly 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 a systematic consequence of batching in high-volume emergency care settings.

4.8. Robustness Checks and Limitations

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 the tendency to batch, 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. Each of these can independently affect LOS and disposition decisions. Removing the potential impact of such unobserved factors might require a fully randomized experiment. However, we find consistency and magnitude in our findings across both 2SLS and UJIVE specifications, with and without precision controls for laboratory ordering and physician characteristics, and across two hospitals with different practice settings. This consistency provides 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 depends on our identification strategy (Section 3.3). To further strengthen our confidence, we performed several robustness checks. First, to address potential many-instrument bias in leniency designs, we implement the Unbiased Jackknife Instrumental Variables Estimator (UJIVE). This estimator uses leave-one-out predictions to eliminate mechanical correlation between the instrument and structural error (Kolesár et al. 2015, Goldsmith-Pinkham et al. 2025). The close agreement between 2SLS and UJIVE estimates across all outcomes (Table 4) increases confidence that many-instrument bias is not driving our findings. Second, to address potential exclusion restriction violations, we conduct sensitivity analyses following Conley et al. (2012). These analyses allow for plausible direct effects of the instrument on outcomes, and our estimates remain significant across a wide range of assumed violation magnitudes (Figure EC.8.1). Third, we conduct a placebo test by applying our BatchTendency instrument (constructed from the main analytical sample) to a placebo sample consisting of rarely batched complaint categories where no imaging tests were ordered. We find null effects on outcomes in this placebo sample (Table EC.7.1), supporting the idea that batch tendency captures batching-specific behavior rather than general physician practice intensity. We acknowledge that we cannot definitively rule out all exclusion restriction violations. We note that to the extent batch tendency captures general physician intensity, our reduced-form estimates still provide the causal effect of being assigned to a high-batching physician.

We also check robustness to other outcome measures. Treatment time, meaning time from doctor contact to disposition (excluding waiting and post-disposition delays), shows no material difference by treatment (Table EC.10.1). For 72-hour returns, regardless of admission, effects remain small and statistically insignificant (Table EC.11.2). For nonlinear models (logit for binary outcomes, Poisson for counts), average marginal effects are similar to linear IV estimates (Table EC.9.1).

To assess whether our results generalize to different clinical scenarios, we conduct heterogeneity analyses across seven chief complaint categories that vary in complexity and prevalence of batching. In every category, we observe the same key patterns: batching increases imaging without efficiency

gains or quality improvements (Figure 5). Our findings are thus not driven by a specific complaint type, which increases confidence in the general applicability of our results.

To address monotonicity, we follow established methods and check whether our instrument has a positive first-stage coefficient across subsamples defined by complaint, severity, capacity, and demographics (Figure EC.6.1). We consistently find stable effect sizes and implications across all robustness checks. Taken together, these robustness analyses give us further confidence in our main conclusion that batching reliably increases imaging and processing delays.

Beyond these methodological considerations, several scope limitations warrant acknowledgment. First, we cannot determine the extent to which the additional imaging from batching represents unnecessary testing versus frontloading of diagnostics that would eventually occur in the inpatient setting. However, interpreting our findings purely as frontloading is complicated by the fact that admission itself is affected by batching. For example, we observe a 39 percentage-point increase in admission probability, suggesting that the tests themselves influence disposition decisions rather than simply shifting testing location. Multiple mechanisms may operate simultaneously. Some additional tests may reveal genuinely admission-worthy conditions, while others may trigger defensive admissions through incidental findings. Distinguishing between these pathways would require linked ED-inpatient imaging data; accordingly, this remains an important direction for future research. To address this concern directly, we re-estimated our main specifications, restricting the sample to patients who were ultimately discharged from the ED. This subsample is informative because, by construction, these patients do not receive subsequent inpatient diagnostic workups, which eliminates the possibility that ED imaging merely substitutes for tests that would have occurred during hospitalization. As shown in Table EC.11.1, the estimated effect of batching on the number of distinct imaging tests remains large, positive, and statistically significant across UJIVE and 2SLS specifications. Magnitudes are comparable to those in the full sample.

Second, our findings apply to moderate-to-high acuity patients with complaints that commonly require multiple imaging studies. This group accounts for approximately 24% of ED encounters and 41% of imaging resource utilization. In contrast, the effects of batching in low-acuity or fast-track settings remain unexplored, as imaging is less common and batching is rare in these contexts. Our sample restrictions, therefore, ensure statistical identification and focus on the clinically meaningful population where batching decisions are consequential and discretionary. As a result, readers should interpret our estimates as applying to this specific margin of physician decision-making.

4.9. Managerial Implications

Our findings have important implications for the management of ED operations. First, while potentially appealing as a workflow efficiency strategy, batch ordering of imaging tests significantly increases ED LOS and resource utilization without corresponding improvements in patient

outcomes. The substantial magnitude of these effects—a 65% increase in LOS and 88% increase in imaging tests for discretionarily batched patients—suggests that ED managers should carefully evaluate policies around physician test ordering. As our decomposition analysis suggests, a meaningful share of these delays operates through increased diagnostic intensity, reinforcing that standard practice serves as a natural filter against unnecessary testing. These findings support Hypotheses 1–3: discretionary batching increases test volume, lengthens processing times, and raises admission probability.

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 in conditions where information from initial tests often eliminates the need for additional imaging. Notably, our heterogeneity analysis shows that the operational inefficiencies of batching persist across all capacity levels—effects are statistically indistinguishable whether the ED is operating normally or under overcapacity (contrary to Hypothesis 4). This suggests that interventions to reduce unnecessary batching are warranted across all levels of capacity utilization, not only during periods of normal capacity.

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. Though preserving physician autonomy in clinical decision making is crucial, our results indicate that full discretion in test ordering timing may lead to suboptimal system performance. Simple interventions, such as providing physicians with feedback on their batching rates relative to peers or implementing decision-support systems that suggest sequential ordering pathways, could help reduce unnecessary testing while maintaining care quality. Our determinants analysis suggests that such interventions may be particularly valuable early in physician shifts, when batching is most prevalent. To better understand the magnitude of these opportunities, we quantify the financial burden of discretionary batching.

Combining our causal estimates with published cost parameters (see EC.5 for details), we adopt a societal perspective using Medicare-allowed reimbursement of $130 per ED radiograph (Dabus et al. 2025) and ED bed-hour costs of $85 based on time-driven activity-based costing studies (Schreyer and Martin 2017, Canellas et al. (2024)). Under deliberately conservative assumptions, discretionary batching imposes at least $349 in direct costs per batched complier: $125 from excess imaging and $224 from excess ED bed utilization. Scaling to annual operations, this amounts to approximately $232,000 per year for a medium-sized ED (40,000 annual visits) and $349,000 for a large urban ED (Table EC.5.1). These estimates represent conservative lower bounds, include only

statistically significant imaging effects (excluding ultrasound and CT point estimates), value bed-hours at operational cost rather than throughput value, and exclude downstream costs from the 40 percentage point increase in admission probability. Inpatient stays cost $10,000–15,000, which would dwarf our direct cost estimates. Notably, we find no offsetting benefits: 72-hour return rates are statistically indistinguishable between strategies. Even under these conservative assumptions, a 50% reduction in discretionary batching would save about $116,000 annually while freeing 850 bed-hours of capacity.

Finally, our findings have implications beyond individual EDs, as increased admission rates from batch ordering create spillover effects throughout the hospital system. Discretionary batching increases admission probability by 40 percentage points, raising patient volumes in inpatient units and potentially inducing speed-up behaviors that can compromise care quality. Hospital administrators should consider these downstream impacts when developing imaging protocols and resource allocation strategies. The consistency of our findings across two institutions with different patient populations and assignment mechanisms (Mayo Clinic and MGH) suggests these patterns are not idiosyncratic to a single setting, strengthening the case for broader policy attention to diagnostic test sequencing in emergency care.

5. Conclusion

Although previous research examines task ordering within centralized protocols, frontline physicians often control when and how they order diagnostic tests, especially in high-pressure settings like EDs. In practice, due to system design or individual choice, doctors self-manage their test-ordering strategies. Because little research addresses the operational impact of test ordering, healthcare managers have limited guidance when physician discretion is involved. Understanding when and how doctors exercise this discretion helps inform system design, encourages better use of discretion, and guides policy to accommodate frontline clinician behaviors.

We address this gap by analyzing the drivers and effects of batch ordering imaging tests in the ED. Using detailed data from two large U.S. EDs—one with random patient-physician assignment—we find that physicians are more likely to batch imaging tests early in their shifts and when the ED is less crowded. This suggests time pressure and ED occupancy measurably influence batching, indicating that physicians adjust orders based on workload and operational demands.

Our results show that batch ordering of imaging tests significantly increases the number of imaging studies performed per patient encounter, thereby increasing resource utilization. We find no evidence that batching reduces patient length of stay (LOS) in the ED or affects 72-hour return admissions. Instead, discretionary batching increases both time spent in the ED and time to disposition, indicating substantial inefficiencies from initiating multiple diagnostic imaging processes

at once. Furthermore, batch ordering substantially increases the probability of hospital admission, with important downstream implications for hospital capacity and patient flow. This supports prior research suggesting sequential test ordering can minimize unnecessary tests by allowing earlier results to inform further decisions. Our findings suggest that while physician discretion can tailor test ordering, it may also promote practices that raise resource use and inefficiency without benefiting patients or the ED system.

Our comparison of batch ordering versus standard practice reflects the real choice facing ED managers: should protocols encourage comprehensive upfront testing or preserve diagnostic flexibility? Our estimates show that preserving optionality through standard practice—which allows information from initial tests to guide subsequent decisions—reduces both testing intensity and processing times without compromising care quality. The consistency of these findings across seven chief complaint categories, from simple presentations like fevers to complex polytrauma cases (Figure 5), simplifies policy recommendations. Interventions to reduce discretionary batching appear warranted across the full spectrum of ED presentations rather than being limited to particular clinical scenarios.

Excess ED testing is not harmless. It increases risks such as unnecessary radiation and both psychological and physical burdens from incidental findings (Müskens et al. 2022). The economic impact is substantial, with test overuse adding to healthcare costs (Atkinson and Saghafian 2023). Our results point to the need to study batching across clinical conditions and settings beyond the ED, considering effects on hospital metrics (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 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 physicians feedback on their ordering patterns and associated outcomes. By aligning physician test-ordering strategies more closely with patient needs (Atkinson and Saghafian 2023), EDs can enhance patient satisfaction and outcomes while improving operational efficiency.

Table 5 Decomposition of Batching Effects on Operational Outcomes

	Log LOS	Log Time to Disposition
Panel A. Product-of-Coefficients Decomposition		
Total Causal Effect (2SLS)		
Batching → Outcome	0.597***	0.651***
	(0.088)	(0.101)
Effects on Mediators (2SLS)		
Batching → Imaging tests ($\hat{\alpha}_1$)	1.241***	1.241***
	(0.116)	(0.116)
Batching → Admission ($\hat{\alpha}_2$)	0.404***	—
	(0.088)	
Controlled Associations: Mediator → Outcome (OLS)		
Imaging tests ($\hat{\beta}_1$)	0.124***	0.034***
	(0.007)	(0.008)
Admission ($\hat{\beta}_2$)	0.042***	—
	(0.009)	
Implied Indirect Effects		
Via imaging ($\hat{\alpha}_1 \times \hat{\beta}_1$)	0.154 (26%)	0.042 (6%)
Via admission ($\hat{\alpha}_2 \times \hat{\beta}_2$)	0.017 (3%)	—
Residual	0.427 (71%)	0.609 (94%)
Panel B. Reduced-Form Attenuation		
Batch Tendency Coefficient		
Baseline	1.046***	1.140***
	(0.135)	(0.159)
+ Imaging tests	0.786***	1.049***
	(0.134)	(0.160)
+ Imaging tests & Admission	0.769***	—
	(0.134)	
Attenuation: + Imaging	24.9%	8.0%
Attenuation: + Imaging & Admission	26.5%	—
Necessary controls	Yes	Yes
Precision controls	Yes	Yes
Observations	11,679	11,679

Notes: Panel A decomposes the total 2SLS effect of batching into components attributable to imaging volume and admission. The IV effects on mediators are from Table 4 (Column 3). Controlled OLS associations condition on the full set of patient severity, demographics, physician characteristics, and temporal fixed effects described in Section 3. Percentages in parentheses represent the share of the total effect. Panel B shows how the reduced-form coefficient on batch tendency attenuates when mediators are added as controls. Standard errors are heteroskedasticity-robust. $^{***}p < 0.001$.

Table 6 Effects of Batching by ED Capacity Status

	Normal Operations (1)	Minor Overcapacity (2)	Major Overcapacity (3)
Log LOS	0.603***	0.645***	0.309
	(0.119)	(0.155)	(0.262)
Log time to disposition	0.638***	0.696***	0.404
	(0.138)	(0.178)	(0.276)
Number of distinct imaging tests	1.201***	1.171***	1.914***
	(0.151)	(0.203)	(0.479)
72-hour return with admission	−0.028	0.006	−0.020
	(0.026)	(0.033)	(0.073)
Batch rate	0.148	0.127	0.125
Observations	5,241	5,355	1,083
Necessary controls	Yes	Yes	Yes
Precision controls	Yes	Yes	Yes

Notes: This table reports two-stage least squares estimates of the impact of batching across different ED capacity levels. ED capacity levels are defined in accordance with internal guidelines in Table EC.2.1. All specifications include time fixed effects and baseline controls. $^{***}p < 0.001$.

Table 7 Determinants of Test Ordering Behavior

	Batched (1)	Number of Imaging Tests (2)
Panel A. Physician Characteristics		
Physician experience (years since residency)	-0.001^{*}	-0.003^{***}
	(0.000)	(0.001)
Physician male	0.006	0.008
	(0.007)	(0.013)
Hours into shift	-0.004^{*}	0.000
	(0.001)	(0.003)
Panel B. ED Conditions		
Capacity Level: Minor overcapacity	-0.017^{**}	-0.026^{*}
	(0.007)	(0.011)
Capacity Level: Major overcapacity	−0.019	-0.017
	(0.011)	(0.019)
Necessary controls	Yes	Yes
Precision controls	Yes	Yes
Observations	11,679	11,679
R^2	0.051	0.121

Notes: Table reports OLS estimates of relationships between physician/ED characteristics and test-ordering behavior. All models include FE for shift and patient characteristics as described in the text. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 8 Comparison of Effects of Batching 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. 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: $^{***}p < 0.001$

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.

Berlin L (2011) The incidentaloma: a medicolegal dilemma. *Radiologic Clinics of North America* 49(2):245–255, ISSN 0033-8389, URL http://dx.doi.org/10.1016/j.rcl.2010.11.002.

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.

Canellas MM, Jewell M, Edwards JL, Olivier D, Jun-O'Connell AH, Reznek MA (2024) Measurement of cost of boarding in the emergency department using time-driven activity-based costing. *Annals of Emergency Medicine* 84(4):410–419, URL http://dx.doi.org/10.1016/j.annemergmed.2024.05.013.

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.

Coen M, Sader J, Junod-Perron N, Audétat MC, Nendaz M (2022) Clinical reasoning in dire times: Analysis of cognitive biases in clinical cases during the covid-19 pandemic. *Internal and Emergency Medicine* 17(4):979–988, ISSN 1828-0447, URL http://dx.doi.org/10.1007/s11739-021-02884-9.

Conley TG, Hansen CB, Rossi PE (2012) Plausibly exogenous. *The Review of Economics and Statistics* 94(1):260–272, URL http://dx.doi.org/10.1162/REST_a_00139.

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.

Dabus G, Hirsch JA, Booker MT, Silva E (2025) Practice expense and its impact on radiology reimbursement. *Journal of the American College of Radiology* URL http://dx.doi.org/10.1016/j.jacr.2025.02.047.

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

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.

Featherston R, Downie LE, Vogel AP, Galvin KL (2020) Decision making biases in the allied health professions: A systematic scoping review. *PLoS ONE* 15(10):e0240716, ISSN 1932-6203, URL http://dx.doi.org/10.1371/journal.pone.0240716.

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.

Goldsmith-Pinkham P, Hull P, Kolesár M (2025) Leniency designs: An operator's manual. URL https://arxiv.org/abs/2511.03572.

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.

Kolesár M, Chetty R, Friedman J, Glaeser E, Imbens GW (2015) Identification and inference with many invalid instruments. *Journal of Business & Economic Statistics* 33(4):474–484, URL http://dx.doi.org/10.1080/07350015.2014.978175.

Kuntz L, Mennicken R, Scholtes S (2014) Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* 61(4):754–771, URL http://dx.doi.org/10.1287/mnsc.2014.1920.

Lam JH, Pickles K, Stanaway FF, Bell KJL (2020) Why clinicians overtest: development of a thematic framework. *BMC Health Services Research* 20(1):1011, ISSN 1472-6963, URL http://dx.doi.org/10.1186/s12913-020-05844-9.

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.

Lumbreras B, Donat L, Hernández-Aguado I (2010) Incidental findings in imaging diagnostic tests: a systematic review. *British Journal of Radiology* 83(988):276–289, ISSN 0007-1285, URL http://dx.doi.org/10.1259/bjr/98067945.

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.

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.

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.

Rao VM, Levin DC (2012) The overuse of diagnostic imaging and the choosing wisely initiative. *Annals of Internal Medicine* 157(8):574–576, ISSN 0003-4819, URL http://dx.doi.org/10.7326/0003-4819-157-8-201210160-00535.

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).

Schreyer KE, Martin R (2017) The economics of an admissions holding unit. *Western Journal of Emergency Medicine* 18(4):553–558, URL http://dx.doi.org/10.5811/westjem.2017.4.32740.

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.

Su H, Meng L, Sangal R, Pinker EJ (2025) Crisis at the core: Examining the ripple effects of critical incidents on emergency department physician productivity and work style. URL http://dx.doi.org/10.2139/ssrn.5113467, available at SSRN: https://ssrn.com/abstract=5113467 or http://dx.doi.org/10.2139/ssrn.5113467.

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, Bartley AC, Smith VD, Didehban R, Lipinski CA, Saghafian S (2016a) Physician in triage versus rotational patient assignment. *The Journal of Emergency Medicine* 50(5):784–790, URL http://dx.doi.org/10.1016/j.jemermed.2016.01.013.

Traub SJ, Stewart CF, Didehban R, Bartley AC, Saghafian S, Smith VD, Silvers SM, LeCheminant R, Lipinski CA (2016b) 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`.