Variation in Batch Ordering of Imaging Tests in the Emergency Department and the Impact on Care Delivery

Jacob Jameson
Harvard Kennedy School, Harvard University

Soroush Saghafian
Harvard Kennedy School, Harvard University

Robert Huckman
Harvard Business School, Harvard University

Nicole Hodgson
Mayo Clinic, Emergency Department

Objectives: To examine heterogeneity in physician batch ordering practices and measure the associations between a physician's tendency to batch order imaging tests on patient outcomes and resource utilization.

Study Setting and Design: In this retrospective study, we used comprehensive EMR data from patients who visited the Mayo Clinic of Arizona Emergency Department (ED) between 10/6/2018 and 12/31/2019. Primary outcomes are patient length of stay (LOS) in the ED, number of diagnostic imaging tests ordered during a patient encounter, and patients' return with admission to the ED within 72 hours. The association between outcomes and physician batch tendency was measured using a multivariable linear regression controlling for various covariates.

Data Sources and Analytic Sample: The Mayo Clinic of Arizona Emergency Department recorded approximately 50,836 visits, all randomly assigned to physicians during the study period. After excluding rare complaints, we were left with an analytical sample of 43,299 patient encounters.

Principal Findings: Findings show having a physician with a batch tendency 1SD greater than the average physician was associated with a 4.5% increase in ED LOS ($p < 0.001$). It was also associated with a 14.8% (0.2 percentage points) decrease in the probability of a 72-hour return with admission ($p < 0.001$), implying that batching may lead to more comprehensive evaluations, reducing the need for short-term revisits. A batch tendency 1SD greater than that of the average physician was also associated with an additional 8 imaging tests ordered per 100 patient encounters ($p < 0.001$), suggesting that batch ordering may be leading to tests that would not have been otherwise ordered had the physician waited for the results from one test before placing their next order.

Conclusions: This study highlights the considerable impact of physicians' diagnostic test ordering strategies on ED efficiency and patient care. The results also highlight the need to develop guidelines to optimize ED test ordering practices.

Key words: Quality of Care/Patient Safety (Measurement); Health Care Costs; Physician Behavior

1. Introduction

Emergency departments (EDs) serve as critical junctures in healthcare delivery, balancing the immediate needs of patients with the overarching operational and administrative demands of hospital management. This balance is precarious and affected by numerous factors, including the strategic ordering of diagnostic imaging tests—a common yet complex practice with implications for patient flow, hospital costs, and patient safety (Cournane et al. 2016). The efficiency of the ED is not just a matter of patient care but also a significant hospital management concern, with the potential to influence hospital-wide operational dynamics and financial health (Darraj et al. 2023, Hodgson et al. (2021)).

An understudied aspect of ED efficiency is the practice of batch ordering imaging tests. Given the long turnaround times of imaging tests, a physician can ensure that their patient is in simultaneous waiting queues for each test by placing multiple orders simultaneously. While ostensibly a measure to expedite patient diagnosis and treatment, batch ordering raises several potential concerns. For instance, the case of a patient presenting with nonspecific abdominal pain could lead to a batch order, including an abdominal CT scan, ultrasound, and X-ray. While comprehensive, this approach raises questions about the necessity of each test, the patient's cumulative radiation exposure, the impact on the patient's length of stay, and overall healthcare costs (Tamburrano et al. 2020, Perotte et al. (2018), Lyu et al. (2017), Jain (2021), Traub et al. (2018b)).

Furthermore, the financial implications extend beyond the cost of the tests themselves. Though sometimes necessary for thorough evaluation, an increased length of stay can also contribute to hospital overcrowding and reduced capacity for new patients, exacerbating operational pressures and financial constraints on the healthcare system (Sartini et al. 2022). This delicate balance between ensuring rapid, accurate diagnosis and minimizing unnecessary use of resources is a central challenge in hospital management, reflecting broader concerns about the sustainability of healthcare practices (Ibanez et al. 2017).

Despite its significance, the impact of batch ordering on these dimensions remains underexplored. The assumption that batch ordering represents an efficient test ordering practice has not been rigorously examined, leaving a gap in our understanding of its true operational and economic implications. This study aims to shed light on this critical issue, exploring how batch ordering of imaging tests affects the length of stay, total testing volume (surrogates for efficiency), and the need for short-term revisits with admission (a surrogate for effectiveness).

By situating this investigation within the context of hospital management, we seek to determine whether the perceived efficiency of batch ordering aligns with its actual outcomes, providing evidence-based insights that can guide future policy and practice in emergency care (Saghafian et al. 2012, Saghafian et al. (2015)).

2. Methods

2.1. Study Design and Setting

Our retrospective observational study was conducted in the Mayo Clinic of Arizona ED. During the study period, the ED recorded 50,836 visits, managed across 26 treatment rooms and up to 9 hallway spaces. The department is exclusively staffed by board-eligible or board-certified emergency physicians (EPs), with rotating residents overseeing about 10% of patient volume. Physicians operate in a unique workflow that includes staggered 8.5-hour shifts and a randomized assignment system that reduces systematic differences in patient populations served by different physicians (Traub et al. 2018a).

We retrospectively reviewed comprehensive ED operational data from 10/6/2018 through 12/31/2019. The dataset includes detailed patient demographics, chief complaints, vital signs, emergency severity index (ESI), length of stay (LOS), and resource utilization metrics. This period was chosen to provide a robust data set while excluding the influence of the coronavirus pandemic. We further restricted our sample to patient encounters serviced by full-time physicians and chief complaint areas seen in over 1,000 encounters over the study period (i.e., excluding rare complaints). Chief complaint categories were created by organizing a patient's "reason for visit" free-text data into broader groupings previously used in the literature (Hodgson et al. 2021). The final sample included 43,299 patient encounters and contained no missing data for covariates used in the analysis.

2.2. Details on Data

A critical aspect of our data is the random patient-to-physician assignment. In most EDs, physicians have some discretion in selecting the patients they see from the pool of those waiting for treatment. In contrast, patients arriving at the Mayo Clinic ED are assigned to physicians via a randomized rotational patient assignment algorithm, which practically removes potential selection bias concerns from our analyses (Traub et al. 2018a). A computer algorithm randomly assigned arriving patients to physicians in a round-robin manner, where assignments were made purely rotationally without considering patient demographics, chief complaint, ESI, physician-patient load, or acuity of patients recently assigned to the physician. In essence, physician-to-patient matching can be deemed random by controlling for patient arrival time and physician shift-level variation. The balance test in Appendix Table 2 confirms that the complaints and severity of patients served are balanced across physicians.

2.3. Definition of Batching

We define "batching" in line with standard emergency medicine practices. Batching occurs when a physician simultaneously orders a comprehensive set of diagnostic tests, typically covering a broad

range of potential diagnoses. This contrasts with sequential ordering, where tests are ordered in sequence based on the information obtained from subsequent tests as needed.

For this study, we focus on batches that include two or more different imaging tests ordered within a 5-minute window at the start of a patient encounter. Sensitivity analyses around this time window, batch size, and when the batch occurs during the patient visit (Appendix Table 3) show that our results are robust to variation in these cutoffs. 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. For our analyses, we focus on batching instances in which the attending physician orders two or more different imaging tests because of the operational implications of scheduling imaging tests that cannot be done in a single scanning session due to differences in equipment and setting.

2.4. Statistical Analysis

To assess the impact of batching on various outcomes of interest, we developed a measure to quantify each physician's overall tendency to batch. While the decision to batch order for a patient itself is endogenous and correlated with both observed and unobserved factors that may impact outcomes, this overall physician "batch tendency" score allows us to utilize an exogenous factor (an instrumental variable as we will describe) to explore the relationship between batching and critical outcomes such as patient length of stay, resource utilization, and 72-hour return to the ED. The batch tendency for each physician was calculated using a "leave-one-out" approach. Specifically, we start by estimating the following multivariable logistic regression:

$$\begin{aligned}\text{logit}(P_{Batched_{i,t}}) = \beta_0 &+ \beta_1 X_{ym} + \beta_2 X_{dt} + \beta_3 X_{complaint \times severity} \\ &+ \beta_4 X_{hypotensive} + \beta_5 X_{tachycardic} + \beta_6 X_{tachypneic} \\ &+ \beta_7 X_{febrile} + \beta_8 X_{physicianID} + \epsilon_{i,t} \end{aligned} \quad (1)$$

Where $Batched_{i,t}$ is a dummy variable equal to one if patient i had their imaging tests batch ordered on the encounter on date t. Covariates include year-month, X_{ym}, to control for time and seasonal variation in batching, such as hospital-specific policies (e.g., initiatives to eliminate excess testing) or seasonality in ED visits. We also control for shift-level variations that include physician scheduling and patient arrival with day-of-week and time-of-day covariates, X_{dt}. Chief complaint by severity, $X_{(complaint \times severity)}$, as well as several other patient-level characteristics such as hypotension, tachycardia, tachypnea, and fever, are included to increase precision and account for variation in patient acuity and clinical presentation. As stated earlier, these controls are required for the patient-to-physician assignment to be deemed as good as random. We use this model to

produce predicted probabilities of batching occurring for each patient encounter on the test sets. We report a 10-fold cross-validated AUC of 0.75 (Appendix Figure 4).

For physician j serving patient i, we then compute the leave-one-out average of $P_{Batched_{(i,t)}}$ for each physician j by excluding the current patient i from the calculation and including all other patients served by physician j during the study period. This measure eliminates the mechanical bias resulting from patient i's own case influencing the physician's batch tendency score and captures the physician's general likelihood of batching imaging tests across a wide range of cases (Eichmeyer and Zhang 2022, Dobbie et al. (2018)).

After calculating each physician's average leave-one-out batch tendency, we center and standardize it into a uniform scale, facilitating more straightforward interpretation and comparison across physicians. Appendix Figure 4 shows that the resulting batch tendency score predicts batch ordering during a specific patient encounter even though it is entirely independent of that patient encounter. The batch tendency score is constructed to reflect physician j's underlying tendency to batch at patient i's encounter and is independent of all patient i's characteristics. Since the decision to batch order tests could be related to patient i's presenting condition, by considering batch tendency as an exogenous instrumental variable, we more robustly estimate the impact of batching, thereby significantly addressing concerns regarding endogeneity (Konetzka et al. 2019).

All statistical analyses were performed using R (version 4.3.2). All multivariable linear regression models control for calendar month and time-of-day fixed effects, which is necessary to achieve quasi-random assignment. We additionally control for patient chief complaint and severity, hypotension, tachycardia, tachypnea, and fever at the time of triage, an indicator for whether any laboratory tests were ordered for the patient, and an ED occupancy measured at the time of patient arrival to improve precision. We use robust standard errors clustered at the physician level.

We evaluate the influence of physicians' batch ordering tendency on three patient-level dependent variables: LOS, 72-hour return with admission, and the number of distinct imaging tests ordered. Additional subgroup analyses explore whether the effect of batching varies across different patient acuities and complaints. Because our data regarding 72-hour returns are limited to returns to the same ED, we expect that the magnitude of our estimate is biased towards the null.

2.5. Data Transformation

As evidenced in the literature, transforming the outcome variable can improve the performance of regression models. For right-skewed outcomes, such as the LOS which is shown to be log-normal, applying a natural log transformation can lead to a more symmetric distribution and mitigate the influence of outliers (Diehr et al. 1999, Cots et al. (2003), Saghafian et al. (2024)). As demonstrated in Appendix Figure 5, the distribution of LOS in our data is highly right-skewed. We thus apply a

natural log transformation to this variable before it is used in our regression analyses. We report the un-exponentiated coefficients from these models in Table 1, which can be interpreted as a $100 \times (e^{\beta} - 1)$ percent change in LOS for a given 1 unit increase in our independent variable of interest, where β is the coefficient on our independent variable of interest.

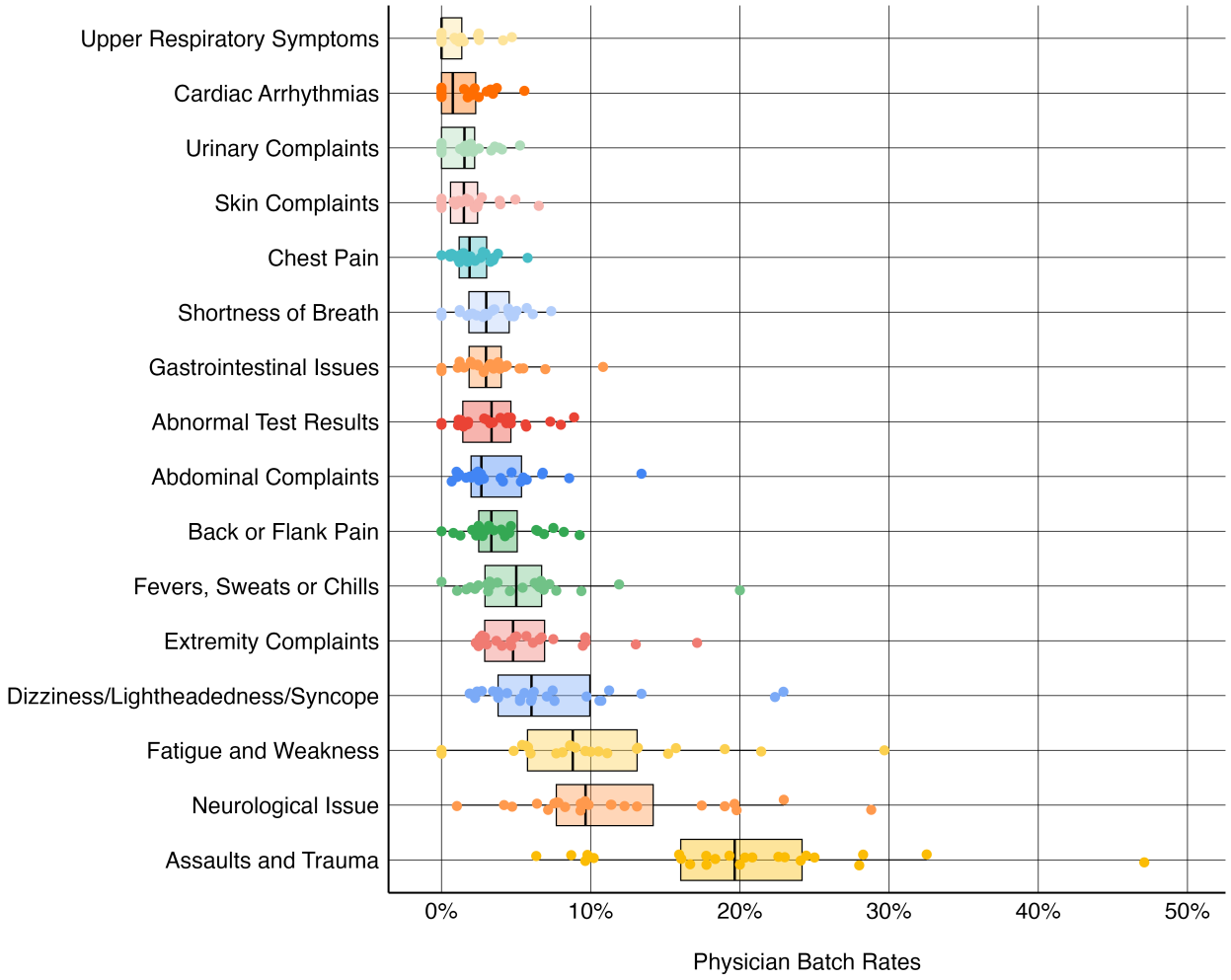
3. Results

The data indicate differences in physician batch ordering practices across complaint categories. Figure 1 displays the crude batch rates calculated for each physician across their patient encounters for each chief complaint. Notably, the variation in batching was most pronounced during patient encounters where the presenting complaint was neurological or trauma-related. We note that at least one imaging test was ordered in 31,498 of the 43,299 patient encounters during the study period. While only 2,421 (7.7%) of these encounters involved image batching, 7,181 (22.8%) of the non-batched encounters resulted in the physician ordering at least one more imaging test after placing the first order.

Table 1 presents the multivariate linear regression coefficients of batch tendency on three primary outcomes: the natural logarithm of ED LOS, 72-hour return and admission, and the number of distinct imaging tests ordered. Our analysis reveals a significant positive association between a physician's tendency to batch order imaging tests and an increased ln(LOS), with a coefficient of 0.044 (95% CI = [0.005, 0.084], $p < 0.001$). This implies that having a physician with a batch tendency 1SD greater than the average physician is associated with a 4.5% increase in ED LOS. However, we also find that a batch tendency 1SD greater than the average physician is associated with a 0.2 percentage point decrease (14.8% decrease) in the probability of a 72-hour return, indicated by a coefficient of -0.002 (95% CI = [-0.003, -0.001], $p < 0.001$), implying that batching may lead to more comprehensive initial evaluations, reducing the need for short-term revisits. Finally, there is a notable association with an increased number of distinct imaging tests ordered, as evidenced by a coefficient of 0.08 (95% CI = [0.065, 0.094], $p < 0.001$). This translates to an additional 8 imaging tests per 100 patient encounters for a physician with a batch tendency 1SD greater than the average physician. Since patients are balanced and randomly assigned to physicians who differ in their batching behavior, this result indicates that batching may lead to tests that would not have been otherwise ordered had the physician waited for the results from one test and used that information to inform the ordering of subsequent tests.

Figure 2 displays the results of the subgroup analysis stratified by the patient's ESI and chief complaint category (as defined in Appendix 3A). We report the standard 95% confidence interval and the Bonferroni-adjusted 95% confidence interval to account for multiple hypothesis testing. Results show heterogeneity in the effect of batch tendency across patient complaints and acuity.

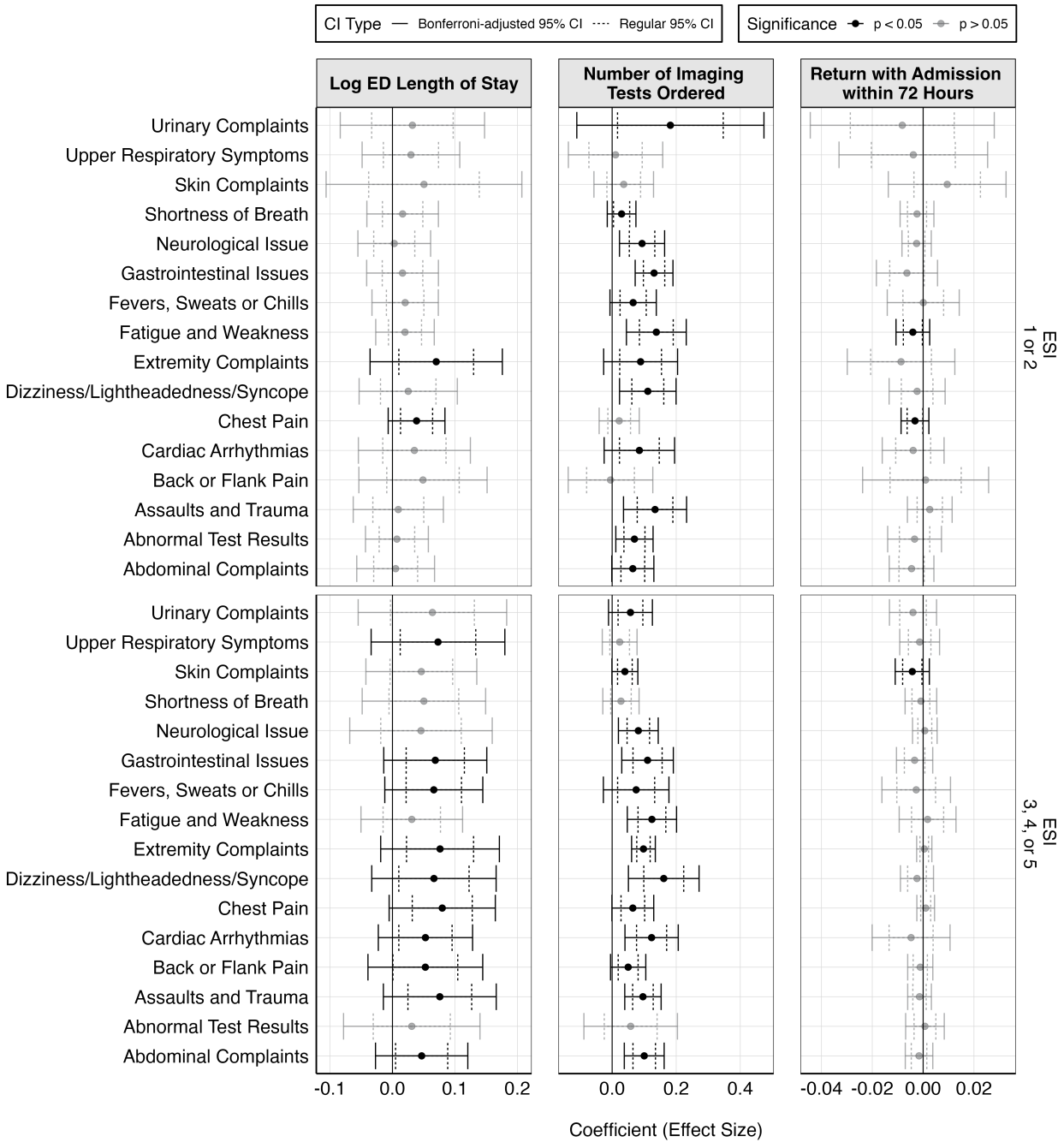
Figure 1 Differences in Imaging Batch Order Frequency by Physicians Across Chief Complaints



This figure highlights the marked differences among physicians in their propensity to batch order imaging tests. Batch rates are crude rates calculated by dividing the number of patient encounters where the physician batch ordered imaging tests for a complaint by the number of patient encounters they had with that complaint. The 24 physicians are represented with points, revealing that specific complaint areas have higher variance than others regarding differing batch rates among physicians.

Notably, among patients of all acuity levels, the propensity to batch order image tests was generally associated with significant increases in the total number of imaging tests ordered. Though not statistically significant at the 0.05 level, the association between batch tendency and LOS and 72-hour return with admission appears to vary in magnitude given the complaint-acuity subgroup.

Figure 2 **Impact of Batch Tendency on Length of Stay, Readmission Rates, and Imaging Utilization for Chief Complaint by Acuity Subgroups**



The coefficient comes from a multivariable linear regression where we regress batch tendency on our primary outcomes for each complaint by acuity subgroup. We control for time and shift fixed effects (necessary for quasi-random assignment), patient-level variables, ED occupancy, whether the patient had laboratory tests ordered during their visit, and vital signs. Standard errors are clustered at the physician level.

Table 1 Multivariable Regression Results of Length of Stay, Readmission Rates, and Imaging Utilization on Batch Tendency

Dependent Variable	Model 1	Model 2	Model 3
Ln(LOS)	0.067***	0.044**	0.044**
	[0.021; 0.112]	[0.005; 0.084]	[0.005; 0.084]
72-Hour Return with Admission	-0.002***	-0.002***	-0.002***
	[-0.003; -0.001]	[-0.003; -0.001]	[-0.003; -0.001]
Number of Distinct Imaging Tests	0.101***	0.080***	0.079***
	[0.083; 0.080]	[0.065; 0.094]	[0.065; 0.094]
Controlled for:			
Patient Arrival and Physician Shift	Yes	Yes	Yes
ED Occupancy	Yes	Yes	Yes
Chief Complaint × ESI	-	Yes	Yes
Labs Ordered	-	Yes	Yes
Vital Signs	-	-	Yes
Observations	43,299	43,299	43,299

The coefficients come from a multivariable linear regression where we regress batch tendency on our primary outcomes. We control for time and shift fixed effects (necessary for quasi-random assignment), patient-level variables, ED occupancy, whether the patient also had laboratory tests ordered during their visit, and vital signs. Standard errors are clustered at the physician level.

$*p < 0.1$; $**p < 0.05$; $***p < 0.01$

4. Discussion

Our study highlights that patterns of diagnostic test ordering in the ED may have implications for the efficiency of care delivery and patient outcomes. Our findings contribute to the growing body of evidence supporting the use of data-driven, personalized approaches in ED management. We can develop more effective, evidence-based strategies for ED resource utilization and patient management by understanding the nuances of test ordering practices and their impact on patient outcomes.

The physician-level variability in inclination towards batching and non-batching test orders—within the same ED environment—raises questions about the underpinnings of clinical decision-making. Notably, our study revealed that physicians with a lower batch tendency, meaning that they employ a more judicious and sequential approach to ordering tests, were associated with a shorter LOS and fewer overall imaging tests. This is due to the information gain advantage of sequential test ordering, where the results of one test may eliminate the need for another. This result aligns with previous research emphasizing the importance of tailored diagnostic pathways

in achieving optimal health outcomes and operational efficacy (Carpenter et al. 2015, Masic et al. (2008), Singh and Graber (2015)).

Over-testing in EDs is not a benign phenomenon. It is associated with increased risks, including patient exposure to unnecessary radiation and the resultant psychological and physical burden from incidental findings (Müskens et al. 2022). Moreover, the economic implications are substantial, with the overuse of diagnostic tests contributing significantly to the escalating costs of health-car(Atkinson and Saghafian 2022). As such, our results suggest the need to examine the practice of batching across different clinical conditions and in other clinical settings beyond the ED (Saghafian and Hopp 2019).

Incorporating physician test ordering tendencies into ED management strategies is complex but potentially beneficial. Recent initiatives have experimented with optimizing patient-physician matching based on various factors, including patient complaints and physician expertise (Saghafian et al. 2018). Our findings suggest that considering physicians' test ordering tendencies, alongside these other factors, could help strike a balance between ensuring thorough patient evaluation and minimizing unnecessary resource utilization. By aligning physician test ordering behaviors more closely with patient needs, EDs may enhance patient satisfaction and outcomes while improving operational efficiency (Saghafian et al. 2018).

Our study involves multiple considerations that may limit the interpretation and application of our findings. While our data involve random assignment of patients to physicians, the variation we observe across physicians could stem from myriad sources, including physician training, accumulated experience, and general inclinations toward more testing (Abaluck et al. 2016). These influences could drive a physician toward a particular testing methodology, confounding the batch tendency measure with other characteristics of the physician's approach to practice. Furthermore, though we consider ED physicians to be independent actors, it is known that they affect each other's speed and quality (Saghafian et al. 2019). Finally, the generalizability of our results may be limited due to the study's single-site design. The Mayo Clinic's operational procedures, patient demographics, and physician culture may not reflect those of other EDs, potentially affecting external validity.

Future studies should investigate the subtleties of the information gain advantage from sequential testing versus the potential benefits of batching. There is a delicate balance between thoroughness and efficiency, which becomes even more precarious in high-stakes environments such as the ED. Understanding and navigating this balance could yield significant advancements in patient care and ED operations.

References

Abaluck J, Agha L, Kabrhel C, Raja A, Venkatesh A (2016) The determinants of productivity in medical testing: Intensity and allocation of care. *American Economic Review* 106(12):3730–3764.

Atkinson M, Saghafian S (2022) Who should see the patient? on deviations from preferred patient-provider assignments in hospitals. *Health Care Management Science* URL `https://ssrn.com/abstract=`, published online.

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

Cots F, Elvira D, Castells X, Sáez M (2003) Relevance of outlier cases in case mix systems and evaluation of trimming methods. *Health Care Management Science* 6(1):27–35, URL `http://dx.doi.org/10.1023/A:1021908220013`.

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`.

Darraj A, Hudays A, Hazazi A, Hobani A, Alghamdi A (2023) The association between emergency department overcrowding and delay in treatment: A systematic review. *Healthcare* 11(3):385, URL `http://dx.doi.org/10.3390/healthcare11030385`.

Diehr P, Yanez D, Ash A, Hornbrook M, Lin D (1999) Methods for analyzing health care utilization and costs. *Annual Review of Public Health* 20(1):125–144, URL `http://dx.doi.org/10.1146/annurev.publhealth.20.1.125`.

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

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.

Hodgson N, Saghafian S, Klanderman M, Urumov A, Traub S (2021) Physician-driven early evaluation: Encounters seen in a vertical model. *JEM Reports* 2(2):100028.

Ibanez M, Clark J, Huckman R, Staats B (2017) Discretionary task ordering: Queue management in radiological services. *Management Science* 64(9):4389–4407.

Jain S (2021) Radiation in medical practice & health effects of radiation: Rationale, risks, and rewards. *Journal of Family Medicine and Primary Care* 10(4):1520–1524, URL `http://dx.doi.org/10.4103/jfmpc.jfmpc_2292_20`.

Konetzka R, Yang F, Werner R (2019) Use of instrumental variables for endogenous treatment at the provider level. *Health Economics* 28(5):710–716, URL `http://dx.doi.org/10.1002/hec.3861`.

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

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

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.

Saghafian S, Austin G, Traub S (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 (2019) Can public reporting cure healthcare? the role of quality transparency in improving patient-provider alignment. *Operations Research (forthcoming)* Published online.

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 W, Van Oyen M, Desmond J, Kronick S (2012) Patient streaming as a mechanism for improving responsiveness in emergency departments. *Operations Research* 60(5):1080–1097.

Saghafian S, Imanirad R, Traub S (2019) Do physicians influence each other's performance? evidence from the emergency department. *Working Paper, Harvard University* Available at https://scholar.harvard.edu/saghafian/publications.

Saghafian S, Kilinc D, Traub S (2024) Dynamic assignment of patients to primary and secondary inpatient units: Is patience a virtue? *Cambridge Handbook on Productivity, Efficiency and Effectiveness in Healthcare (forthcoming)* (available on SSRN).

Sartini M, Carbone A, Demartini A, et al. (2022) Overcrowding in emergency department: Causes, consequences, and solutions-a narrative review. *Healthcare* 10(9):1625, URL http://dx.doi.org/10.3390/healthcare10091625.

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

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. (2018a) The durability of operational improvements with rotational patient assignment. *American Journal of Emergency Medicine* 36(8):1367–1371.

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

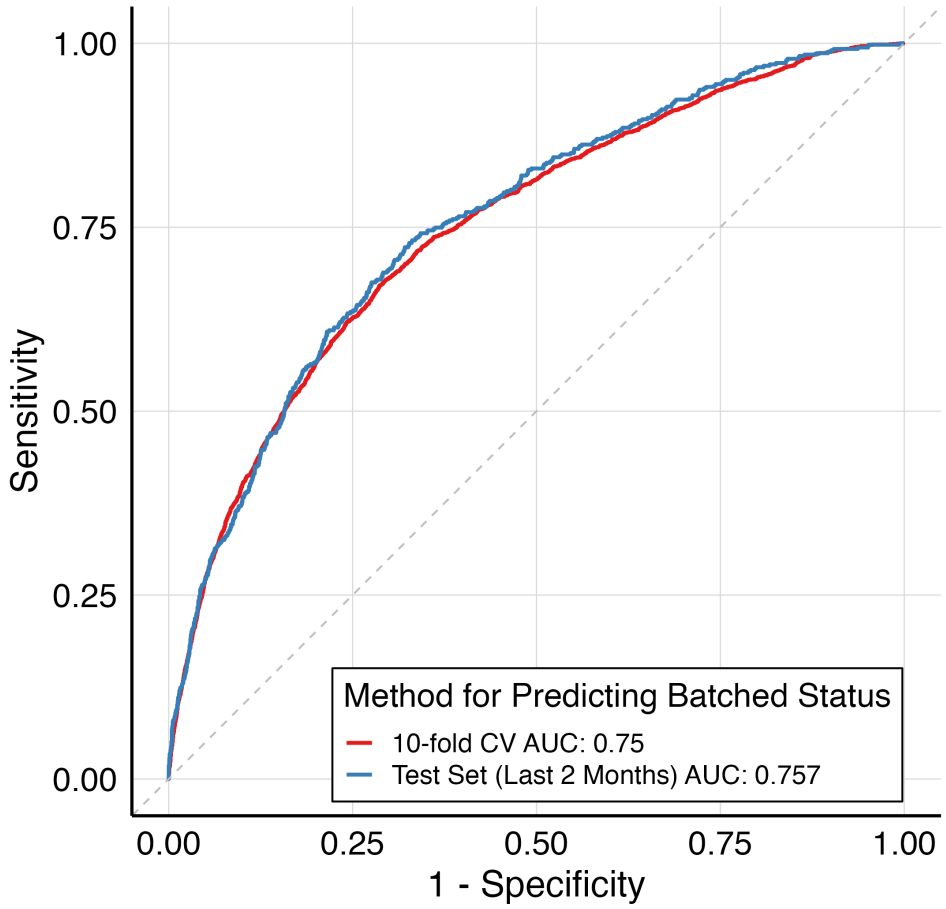
Appendix. General appendix

Table 2 1A: Wald Test Results for Equivalence of Chief Complaints, Emergency Severity, and Vital Signs Across Physicians

Chief Complaints	**Frequency (%)**	**F-Statistic**	**p-value**	**Adj. p-value**
Abdominal Complaints	6,232 (14%)	1.401	0.095	1.00
Abnormal Test Results	1,816 (4%)	2.015	0.002	0.07
Back or Flank Pain	2,550 (6%)	1.029	0.423	1.00
Cardiac Arrhythmias	1,044 (2%)	0.912	0.583	1.00
Chest Pain	3,521 (8%)	1.042	0.406	1.00
Dizziness/Lightheadedness/Syncope	1,894 (4%)	0.921	0.570	1.00
Extremity Complaints	5,259 (12%)	0.991	0.472	1.00
Assaults and Trauma	2,381 (5%)	0.773	0.769	1.00
Fatigue and Weakness	1,494 (3%)	0.681	0.869	1.00
Fevers, Sweats, or Chills	1,842 (4%)	1.132	0.299	1.00
Gastrointestinal Issues	3,323 (8%)	1.027	0.425	1.00
Neurological Issue	3,492 (8%)	0.707	0.843	1.00
Shortness of Breath	2,962 (7%)	1.198	0.232	1.00
Skin Complaints	2,176 (5%)	1.021	0.433	1.00
Upper Respiratory Symptoms	1,915 (4%)	1.239	0.197	1.00
Urinary Complaints	1,399 (1%)	1.837	0.009	0.213
Emergency Severity Index (ESI)	**Frequency No. (%)**	**F-Statistic**	**p-value**	**Adj. p-value**
ESI 1	450 (1%)	0.884	0.621	1.00
ESI 2	13,463 (31%)	1.304	0.149	1.00
ESI 3	24,679 (57%)	0.867	0.645	1.00
ESI 4	4,572 (11%)	1.414	0.090	1.00
ESI 5	135 (0%)	1.302	0.151	1.00
Vital Signs	**Frequency No. (%)**	**F-Statistic**	**p-value**	**Adj. p-value**
Tachycardic	8,360 (19%)	1.390	0.100	1.00
Tachypneic	3,996 (9%)	1.176	0.254	1.00
Febrile	1,021 (2%)	1.708	0.019	0.463
Hypotensive	645 (1%)	1.054	0.390	1.00

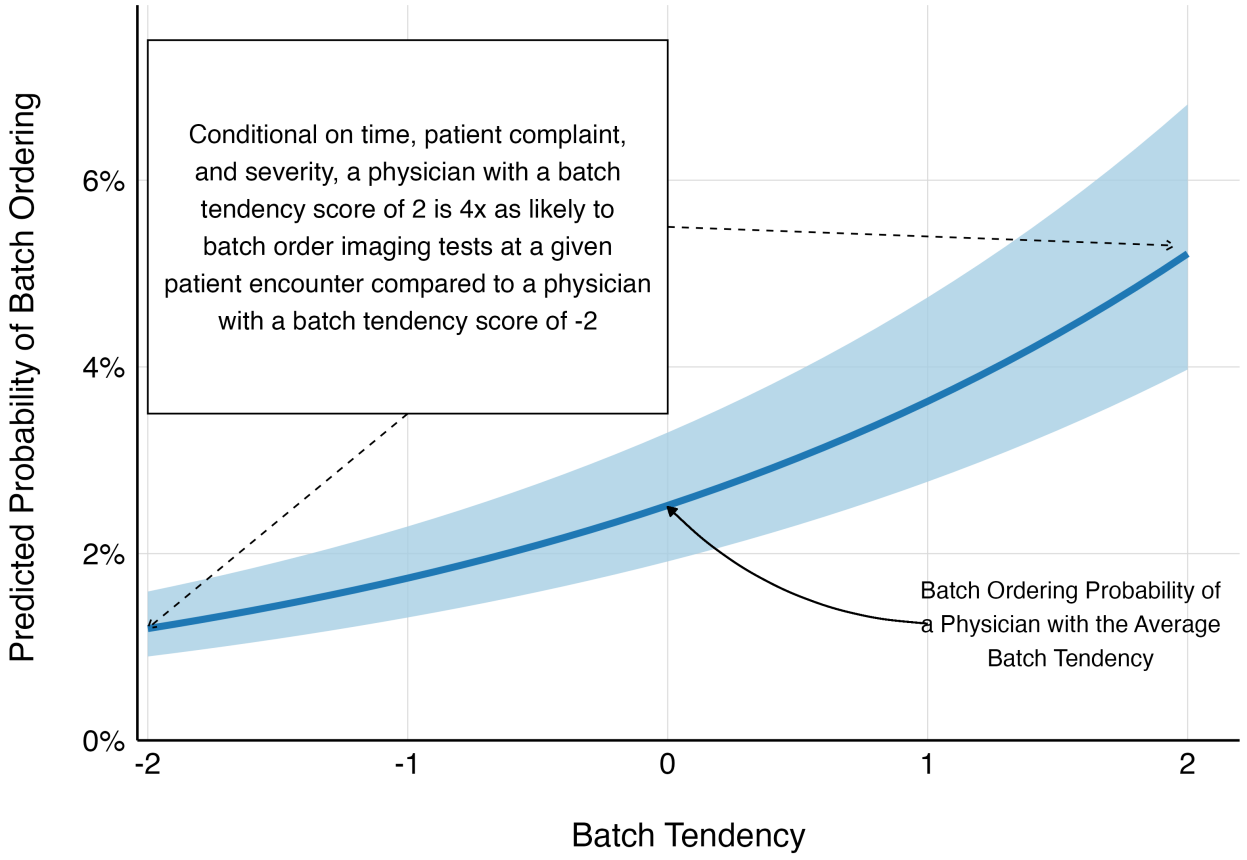
This table presents the results of Wald tests designed to evaluate the balance of chief complaints, Emergency Severity Index (ESI), and patient vital signs across physicians in our dataset. The tests compare models with and without physician identifiers as predictors for each listed variable, assessing whether the distribution of these variables is consistent across physicians, as expected under random assignment. Each row represents a separate Wald test, with the F-statistic indicating the test strength and the p-value showing the likelihood that any observed differences could occur by chance. A balanced distribution, indicated by non-significant p-values, is expected.

Figure 3 1A: ROC Curves and AUC Scores for Logistic Regression Predicting Batching Status



This figure presents the ROC Curves and AUC scores for logistic regression models predicting whether batching will occur for a given patient encounter. We compare the test AUC of a model that uses the last 2 months of data as our test set with a model that uses 10-fold cross-validation.

Figure 4 **2A: Impact of Physician Batch Tendency on Predicted Batch Ordering Probability.**



This figure shows the effect of a physician's "batch tendency" on the predicted probability (with 95% confidence intervals) of batch ordering at a given patient encounter, conditional on time, patient complaint, and severity, from a logistic regression model controlling for these features. The construction of the batch tendency measure is described in the manuscript. Batch tendency, which is entirely independent of the given patient encounter, is predictive of batching during the given encounter. Standard errors are clustered at the physician level.

Figure 5 **3A: Distribution of ED Length of Stay Before and After Log Transformation**

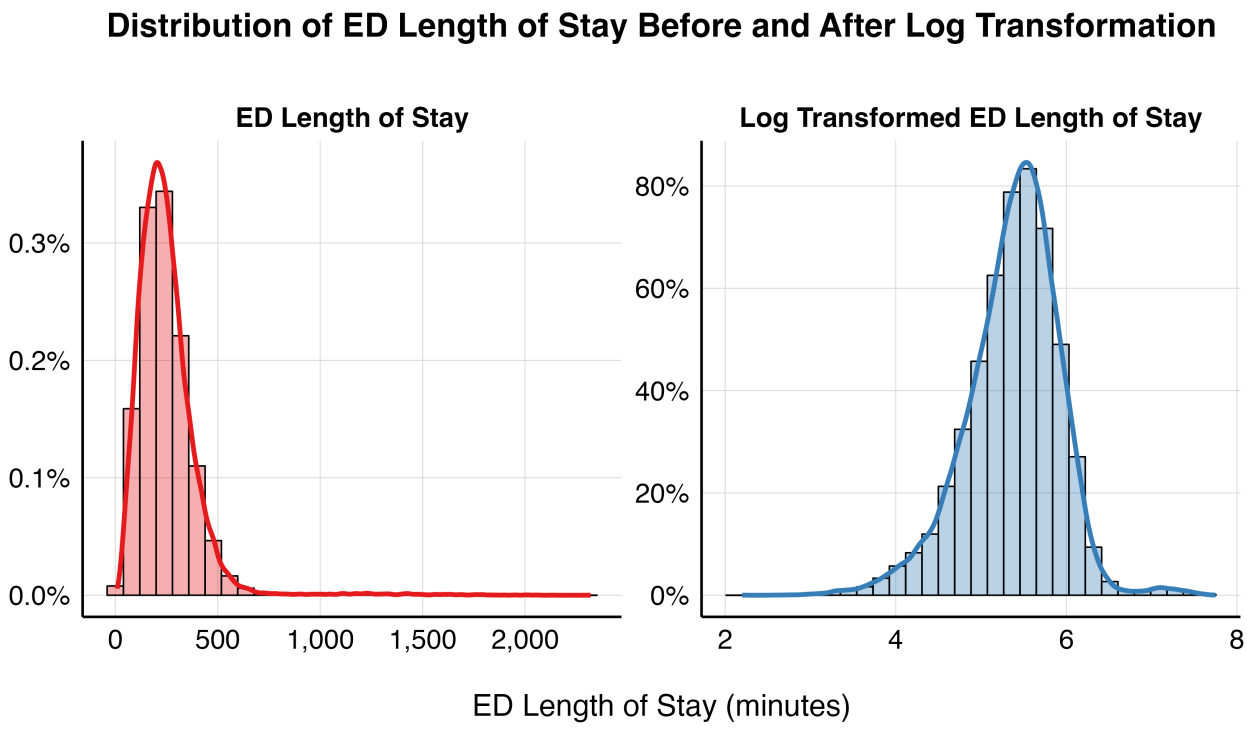


Table 3 2A: Robustness of Batching Definition

Criteria	**ln(LOS)**	**72HR RA**	**NUM IMAGES**
Batching Criteria is 1 minute	0.044**	-0.002***	0.079***
	(0.020)	(0.001)	(0.007)
Batching criteria is 10 minutes	0.042**	-0.001***	0.080***
	(0.020)	(0.0005)	(0.007)
Batching criteria is 20 minutes	0.039**	-0.002***	0.081***
	(0.020)	(0.0004)	(0.006)
Batching criteria is 30 minutes	0.038*	-0.002***	0.083***
	(0.021)	(0.0004)	(0.005)
Batching criteria is $\geq$ 3 tests	0.012	-0.002***	0.047**
	(0.020)	(0.001)	(0.023)
Batching criteria is $\geq$ 4 tests	-3.190	-0.073	3.994*
	(3.175)	(0.086)	(2.165)
Includes batches not done as first set of tests	0.046**	-0.002***	0.080***
	(0.021)	(0.0005)	(0.007)

The coefficient comes from a multivariable linear regression where we regress batch tendency on our primary outcomes. We control for time and shift fixed effects (necessary for quasi-random assignment), patient-level variables, hospital occupancy, whether the patient also had laboratory tests ordered during their visit, and vital signs. The table shows that results are robust to batching definition. Standard errors are clustered at the physician level.

$*p < 0.1$; $**p < 0.05$; $***p < 0.01$