To Batch or Not to Batch: Test-Ordering Variabilityin the ED

Jacob Jameson
Interfaculty Initiative in Health Policy, Harvard University, jacobjameson@g.harvard.edu

Soroush Saghafian
Kennedy School of Government, Harvard University, soroush_saghafian@hks.harvard.edu

Robert Huckman
Harvard Business School, rhuckman@hbs.edu

Nicole Hodgson
Mayo Clinic, Hodgson.Nicole@mayo.edu

We use practice variation across physicians to uncover the role of test-ordering on care delivery in the ED. Using records of over 45,000 Mayo Clinic emergency department visits, we find that quasi-random assignment to a top (versus bottom) decile batching physician significantly increases excess diagnostic testing. Instrumental variable results show that batching leads to an additional 13.5 tests per 100 batches, resulting in excess tests that would not have been ordered under a sequential-ordering strategy. Contrary to expectations, batch ordering does not reduce ED length of stay or affect 72-hour return rates, challenging the perceived efficiency of batching and highlighting the necessity for more targeted testing strategies in emergency care.

Key words: Emergency Department, Operational Effeciency, Diagnostic Testing

---

1. Introduction

Healthcare delivery, particularly in the emergency department (ED), is a delicate balance that involves ensuring optimal patient outcomes while optimizing resource utilization. Achieving these twin goals requires timely and accurate diagnosis, which in turn enables prompt and appropriate treatment, consequently improving patient prognosis and reducing the likelihood of adverse events. Furthermore, efficient patient discharge from the ED can help alleviate overcrowding, a severe issue with potential consequences including higher complication rates and increased mortality Bernstein et al. (2009).

One important factor that can impact the speed and effectiveness of diagnosis in the ED is the availability and performance of diagnostic tests Balogh et al. (2015). A variety of diagnostic tests are used in the ED, including laboratory tests, imaging studies, and specialized tests. These tests can provide valuable information about a patient's condition and help to guide treatment decisions.

A critical question in this context pertains to whether physicians in the ED should batch order diagnostic tests or order them sequentially. This decision essentially represents a tradeoff between reducing patient length of stay and risk of over-testing. Over-testing, or performing unnecessary tests, can lead to increased costs, unnecessary patient anxiety, and potential harm from follow-up of false-positive results Koch et al. (2018). Conversely, keeping a patient for an extended time to perform all possible tests could lead to ED overcrowding, an issue associated with severe consequences, as mentioned earlier. Instead, what is needed is a reasonable balance between the number of diagnostic tests performed and the total time the patient is kept in the ED before either being admitted or discharged. Several studies have demonstrated that optimizing the ED patient flow process can result in significant improvements Saghafian et al. (2015), however, research surrounding test ordering strategies to improve the patient flow processes remains limited.

In this paper, we use data from over 45,000 patient visits to the ED that occur during our study period to quantify the benefits and consequences of batching versus sequentially ordering diagnostic tests on patient length of stay, re-admission, and resource utilization. Our empirical strategy exploits random assignment of patients to ED physicians who differ in their propensity to batch-order diagnostic tests. When patients arrive at the ED, they are assigned to a physician based on availability, with no discretion on either side. Thus, patients who arrive at the ED at similar times are randomly assigned to physicians who vary in their willingness to batch order diagnostic tests. We measure physician tendency to batch using a leave-out, residualized measure based on all other patients the physician has seen in the ED in the study period. The tendency measure strongly predicts the ED test batch outcome but is uncorrelated with patient and ED visit characteristics.

We start by evaluating the reduced-form effects of ED provider batch-ordering tendency on downstream patient outcomes and turnaround time. We find that practice variation as captured by physician batch-ordering tendency has large and significant consequences. Being treated by a provider in the top decile of the tendency distribution, compared to being treated by someone in the bottom decile results in an additional 2 diagnostic test per 100 patient encounters after controlling for the physician's underlying propensity to test.

Physician care takes on multiple dimensions besides the decision to batch order diagnostic tests. Employing a placebo exercise, we find evidence that that excess testing occurs after having seen a high-batching physician is due to batching from this physician, as opposed to other differences across physician in care that correlate with batching tendency: we find no effects of physician batch tendency on key outcomes for a "placebo sample" of patients who visit the ED for conditions that rarely result in batched orders.

With the caveat that other unobserved dimensions of physician care may impact patient outcomes, we employ physician batch tendency as an instrumental variable for having tests batch ordered in the ED to quantify the effect of batching directly. Because of the institutional features of the ED, our research design closely approximates an RCT that assigns patients to batch-ordering or sequential-ordering arm. In the ED, patients have no discretion over choosing providers, and in our specific ED, physicians have discretion over choosing patients, alleviating major selection issues present in other health care settings. Furthermore, physicians exhibit wide variation in practice behavior in batch-ordering, even within the same hospital, while following the same guidelines. Finally, patient-physician interactions in the ED are typically well documented, short, and one-off, constraining physician decision-making to a more limited, better-observed choice set than present in settings such as specialty or primary care.

In sum, exploiting practice variation in ED settings shuts down other (but not all) potential channels besides test batching that are present in other settings, determine length of stay, and impact patient outcomes. This approach allows us to move closer to identifying the causal impact of batch-ordering diagnostic tests on patient outcomes and resource utilization. It is important to note that this paper studies the impact of batch-ordering through a batching decision requiring clinical judgment (within practice norms) rather than through specific hospital policies, differences in adherence to clinical practice guidelines, or substandard care.

The remainder of this paper is structured as follows. The next section describes the data source and outlines our baseline sample. The empirical strategy and its accompanying identifying assumptions are laid out in Section III. Section IV presents the results. Section V draws implications for batch-ordering policies. The last section concludes.

2. Data and Definitions

Our study was conducted in the Emergency Department (ED) of the Mayo Clinic of Arizona, a tertiary care hospital without obstetrical services, an inpatient pediatrics unit, or a trauma designation. During the study period, the ED recorded approximately 43,000 visits per year, managed across 26 treatment rooms and up to 9 hallway spaces. The department is exclusively staffed by board-eligible or board-certified emergency physicians (EPs), with rotating residents overseeing about 10% of patient volume. Physicians operate in a unique workflow that includes staggered 8.5-hour shifts and a rotational patient assignment system, minimizing potential selection bias in patient encounters.

We conducted a retrospective review of comprehensive ED operational data from 10/6/2018 through 12/31/2019, coinciding with the initiation of a new electronic medical record. 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.

2.1. Sample Construction

Our research design focuses on adults who visit the Mayo Clinic of Arizona ED. We observe approximately $48,000$ such visits during the study period. To improve power, we drop encounters with rare chief complaints (< 1000 total encounters of this kind) and complaints where a batch order occurs less than 10 percent of the time. Since batch orders are rare for these cases, our physician batch tendency instrument could suffer from a weak instrument problem if we included them. Examples of complaints dropped include urinary complaints and allergic reactions. Excluding these conditions does not introduce selection bias unless physician test batching tendency is orthogonal to physician diagnosing behavior. While this assumption may be violated if we were to use a very detailed level of chief complaint information upon which to base our exclusion criterion, it is plausibly satisfied when using broad complaint categories. In order to estimate a precise measure of physician-level batch tendency, we further restrict our sample to the $42,000$ encounters involving full-time physicians who treat over 500 ED cases per year.

2.2. Variable Definitions

Our explanatory variable in the IV analysis, $Batched_i$, is an indicator for whether patient i has their tests batch-ordered at their ED encounter. While the patient could decide not to undergo the tests ordered by the physician, this is rare in practice. Below, we detail our primary outcomes: (a) ED length of stay (LOS), and (b) resource utilization.

(a) ED length of stay (LOS).– ED-LOS is a critical measure of efficiency and patient throughput in emergency care settings. It is defined as the duration from a patient's arrival to the ED until their departure, whether by discharge or admission to the hospital. The hypothesis is that batch testing may lead to shorter ED-LOS, potentially improving patient throughput and reducing crowding.

(b) Resource Utilization– Resource utilization in the ED typically refers to the extent of medical services and interventions a patient receives. In this study, we quantify resource utilization by the total number of distinct diagnostic tests ordered per patient during their ED stay. This encompasses both initial and subsequent tests. The hypothesis is that batch testing may lead to variations in the number of tests ordered, potentially influencing the overall healthcare expenditure and efficiency.

2.2.1. Batching We define "batching" in line with standard emergency medicine practices. It involves a physician ordering a comprehensive set of diagnostic tests at once, typically covering a broad range of potential diagnoses. This contrasts with non-batching, where tests are ordered more selectively based on the information available at the time, with additional tests potentially ordered later as needed.

We operationalize batching as the ordering of multiple diagnostic tests within a 5-minute window. In Section 4.4, sensitivity analyses on this cutoff point showed that our results are robust to this definition. This operational definition aligns with the concept of batching as a single comprehensive effort. For our analysis, all lab-based testing ordered in this timeframe are categorized as one distinct type of testing (lab). In contrast, each imaging test (e.g., X-ray, CT scan) are considered a separate distinct test. Therefore, a batch in our study is defined as consisting of two or more distinct diagnostic tests, which could be a combination of lab and imaging tests or multiple imaging tests.

This approach reflects the realistic diagnostic strategies in an ED setting and allows for a clear distinction between batching and non-batching behaviors. The differentiation between lab tests as one collective unit and imaging tests as individual units is based on their operational impact in the ED, particularly regarding patient waiting time and processing queues.

2.3. Summary Statistics

Table 1 presents a detailed overview of the emergency department (ED) characteristics, patient demographics, and medical tests for our baseline sample, derived from the data collected during the study period. Data is at the patient encounter level. On average, the emergency department manages a volume of approximately 24 patients (Mean = 24.18), indicating a significant but manageable patient load. Key physiological markers such as tachycardia (Mean = 0.19) and tachypnea (Mean = 0.09) are prevalent among patients, albeit at varying degrees, highlighting critical aspects of emergency care. Notably, a small proportion of patients exhibit febrile (Mean = 0.02) and hypotensive (Mean = 0.01) conditions, emphasizing the diversity of cases encountered in the ED setting. The Emergency Severity Index (ESI) averages at 2.78, suggesting that most cases fall within the moderate acuity level.

Common chief complaints in the emergency department paint a vivid picture of patient needs and healthcare demands. The data reveals abdominal complaints (Mean = 0.14), extremity complaints (Mean = 0.12), and chest pain (Mean = 0.08) as the most frequently occurring, followed by neurological (Mean = 0.08) and gastrointestinal issues (Mean = 0.08). This information is crucial in understanding the primary reasons for ED visits and the required resources for patient care.

The patient characteristics in our sample depict a diverse demographic landscape. The average arrival age in the ED is approximately 58 years (Mean = 58.33), indicating a predominantly middle-aged population. The racial composition is predominantly White (Mean = 0.89), with smaller proportions of Black (Mean = 0.04) and Asian (Mean = 0.03) patients. Notably, a slight majority of the patients are female (Mean = 0.54), offering insights into gender dynamics in healthcare utilization.

In terms of diagnostic tests, the data reveals a high reliance on lab tests (Mean = 0.77), indicating their crucial role in patient diagnosis and management. X-Ray (Mean = 0.47) and CT scans, both non-contrast (Mean = 0.21) and contrast (Mean = 0.19), are also frequently employed, underscoring the importance of imaging in modern emergency medicine. Ultrasound usage (Mean = 0.11) and the practice of batch ordering tests (Mean = 0.38) further reflect the operational aspects of the ED and the strategies employed to manage patient care effectively.

Table 1 Summary Statistics of Emergency Department Encounters

	Mean	Q1	Median	Q3
ED Characteristics				
ED Volume	24.18	15	25	32
Tachycardic	0.19			
Tachypneic	0.09			
Febrile	0.02			
Hypotensive	0.01			
ESI	2.78	2	3	3
Complaint: Abdominal	0.14			
Complaint: Extremity	0.12			
Complaint: Chest Pain	0.08			
Complaint: Neurological	0.08			
Complaint: Gastrointestinal	0.08			
Patient Characteristics				
Arrival Age	58.33	44	62	75
Race: White	0.89			
Race: Black	0.04			
Race: Asian	0.03			
Gender: Female	0.54			
Tests				
X-Ray	0.47			
Ultrasound	0.11			
Non-Contrast CT	0.21			
Contrast CT	0.19			
Lab	0.77			
Tests were Batch Ordered	0.38			

This table reports summary statistics for the baseline sample of emergency department visits during the study period described in the text. 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*).*

3. Empirical Strategy

Our empirical strategy closely follows the literature that relies on quasi-random assignment of agents to cases, often referred to as the "judges design." Papers in this literature typically exploit variation in the sentencing leniency of judges who work in the same court. Similarly, we explore batching variation across physicians who work in the same emergency department. In its reduced form, under the assumption of quasi-random assignment, this approach allows researchers to identify the causal effect of being assigned to different types of physicians. Under additional assumptions, an instrumental variable approach identifies the causal effect of a given medical decision. We employ both approaches and lay out their details in the next subsections.

3.1. Institutional Details on Patient-Physician Assignment

Contrary to most healthcare settings where patients exhibit choice, they are predominantly passive in their physician assignment in the ED. In most EDs, however, physicians have discretion in picking their patients. In contrast, patients arriving at the Mayo Clinic ED are randomly assigned to physicians via a rotational patient assignment algorithm Traub et al. (2016), which removes potential selection bias concerns for our analyses. In essence, barring arrival time and shift-level variation, the physician-to-patient matching can be deemed random. Table 2 displays that patient encounters (regarding chief complaints and emergency severity) are equitably distributed across physicians within our study's cohort.

Table 2 Balancing Test: Wald Test for Equality of Means

Chief Complaints	Frequency	F-Statistic	$Pr(>F)$
Abdominal Complaints	6232	2.587	0.108
Back or Flank Pain	2552	1.637	0.201
Chest Pain	3525	0.407	0.524
Extremity Complaints	5265	1.847	0.174
Falls, Motor Vehicle Crashes, Assaults, and Trauma	2381	0.023	0.880
Gastrointestinal Issues	3323	0.105	0.746
Neurological Issue	3495	0.135	0.713
Shortness of Breath	2966	1.324	0.250
Skin Complaints	2178	0.383	0.536
Upper Respiratory Symptoms	1917	0.017	0.896
Emergency Severity	**Frequency**	**F-Statistic**	$Pr(>F)$
ESI 1 or 2	13914	0.011	0.915
ESI 3, 4, or 5	29386	0.010	0.921

Table 2 reports the results of a Wald test which was conducted to assess the balance of chief complaints across providers in our dataset. A balanced distribution implies that complaints and severity are evenly distributed across providers, which we expect to be the case due to randomization. The Wald F-statistic and p-value are reported. Robust standard errors (type HC1) were used to account for potential heteroscedasticity in the data.

3.2. Batch Tendency Construction

To measure physician batch tendency, we use the physician's residualized leave-out average batch rate. This measure is derived from two steps following the approaches taken by 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.

$$Batched_{i,t} = \alpha_0 + \alpha_{ym} + \alpha_{dt} + \alpha_{complaint\_esi} + \varepsilon_{i,t} \tag{1}$$

Where $Batched_{i,t}$ is a dummy variable equal to one if patient i had their diagnostic tests batched on encounter that took place on date t. Fixed effects include year-month fixed effects, α_{ym}, to control for time and seasonal variation in batching, 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 both physician scheduling and patient arrival with day of week-time of day fixed effects, α_{dt}. Chief complaint by severity fixed effects, $\alpha_{complaint}$, were also included to increase precision. As stated earlier, these controls are what is required for our quasi-random assignment assumption. Under the assumption that we have captured the observables under which quasi-random assignment occurs in the ED, the unexplained variation– the physician's contribution– resides in the error term, $\varepsilon_{i,t}$.

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

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

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

We document that the Mayo Clinic ED physicians exhibit wide, systematic variation in their propensity to batch order diagnostic tests. Figure 1 graphs the histogram of batch-ordering frequency by physician for popular chief complaints, highlighting that the variation in batching differs systematically. Table 3 presents the "first stage" in a regression table: being assigned to a 10pp higher batch-tendency physician is associated with a 6pp increase in the likelihood of having tests batch-ordered in the ED. The F-statistic is 94 when all controls and fixed effects are included.

The coefficient is greater than one because all emergency visits are used to construct the tendency instrument, while the first stage is calculated using the baseline sample only, which excludes the rare complaints.

Figure 1 Batch Frequency by Physician for High Volume Complaints

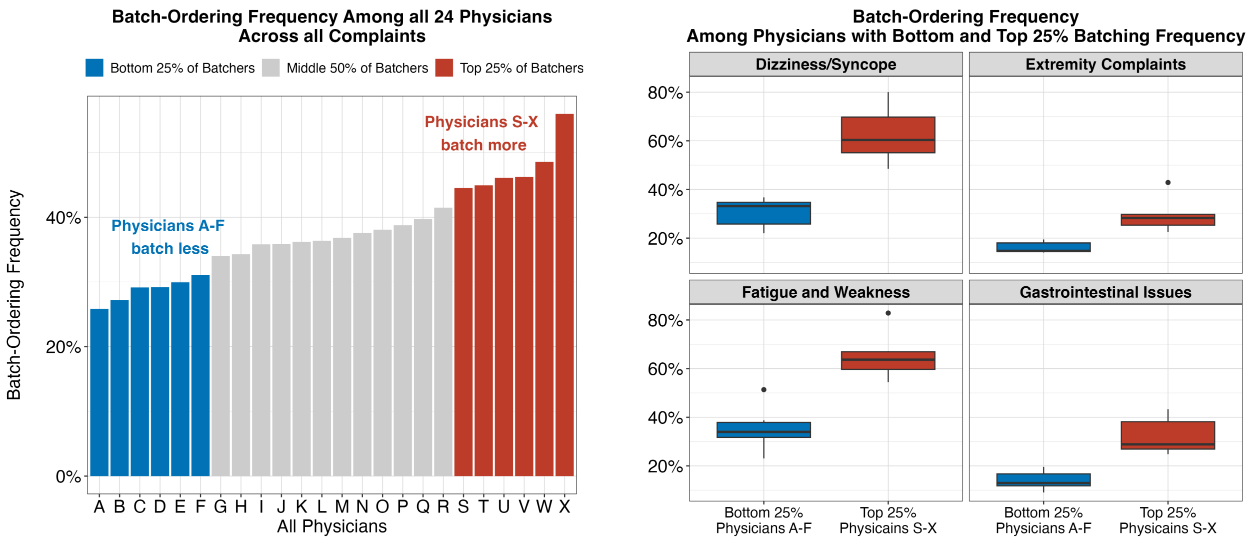


Figure 1 illuminates the marked differences among physicians in their propensity to batch-order diagnostic tests. Physicians are mapped on the x-axis, revealing those with a systematically heightened tendency to batch (in red) compared to their peers who batch less frequently (in blue). The quartile of batchers is calculated based on the batching rate across all complaint areas per physician. Complaints presented in this figure are from patient encounters that exhibit the highest variance in physician batching. Physicians who batch (not batch) in one category of complaints tend to also batch (not batch) in other categories.

To estimate the reduced-form effects of being treated by a batch-preferring physician, we estimate the following equation:

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

This reduced form will allow us to check that our instrument is a strong instrument. To study the effects of test batching in the ED on an outcome Y_i, we estimate the following 2SLS equations using our baseline sample:

$$Y_i = \beta_0 + \beta_1 Batched_i + \theta X_i + \varepsilon_i \tag{4}$$

$$Batched_i = \delta_0 + \delta_1 Tendency_{i,j}^{phys} + \delta_2 X_i + \nu_i \tag{5}$$

Table 3 First Stage: Effect of Batch Tendency on Test Batching

	Model 1	Model 2	Model 3
Batching Tendency	1.06 ***	1.06 ***	1.06 ***
	(0.01)	(0.01)	(0.01)
F statistic (full model)	23.87	94.41	90.44
F (full model): p-value	< 0.001	< 0.001	< 0.001
Num. obs.	41929	41929	41929
Seasonality and shift fixed effects?	Yes	Yes	Yes
Chief Complaint?	No	Yes	Yes
Patient observables?	No	No	Yes

Estimates of the first stage for the baseline sample described in the text. Seasonality shift fixed effects include Year-Month and Hospital-Day of week-Hour of day fixed effects. Chief complaint comes from the cleaned complaint that the patient came in with at the initial encounter. Patient observables include sex dummy, race/ethnicity, and age bins. Column 3 corresponds to the baseline controls. Robust standard errors are clustered at the physician level.

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

Where Y_i represents our main outcomes of interest: length of stay, 72 hour readmission, and resource utilization. and X_i is the same as in the reduced-form approach. $Batched_i$ variable suffers from potential endogeneity concerns. For example, injury severity may be unobserved and correlated with need to run multiple tests, which in turn also affects length of stay. Hence, we instrument $Batched_i$ with the assigned physician j's underlying tendency to batch, $Tendency_{i,j}^{phys}$. We cluster robust standard errors at the physician level to account for the assignment process of patients to physicians.

3.3. Identifying Assumptions

The reduced-form approach delivers an unbiased estimate of the causal effect of being treated by a higher tendency to batch physician, since assignment of patients to ED physicians is random, conditional on seasonality and shift ("conditional independence"). The residualization in equation (1) controls for more controls than required to achieve quasi-random assignment; they are included for statistical precision in measuring physician tendency to batch.

Our instrumental variable approach, which aims to recover the causal effect of having diagnostic tests batch ordered, relies on three additional assumptions: relevance, exclusion, and monotonicity. We reported a strong first stage (i.e., relevance) at the end of the previous Section. The exclusion restriction requires that the instrument must influence the outcome of interest only through its effect on test batching. This is perhaps our strongest assumption and is at its core, untestable. However, several features of the ED setting suggest that such violation may likely only have a small impact and may be less concerning than in other health care settings. First, unlike in primary care settings, where the patient and primary care provider have many repeat encounters, the scope of what the emergency physician can do to impact medium-term outcomes is limited and well-observed by the researcher. Second, any violation of the exclusion restriction needs to directly affect the specific outcome of interest. The channel by which ED physicians can influence length of stay relative outcomes is likely through testing and diagnosis. Nevertheless, we take this assumption seriously and perform a placebo check in Section 4.2 as well as various robustness checks in Section 4.4.

Finally, the monotonicity assumption is necessary for interpreting the coefficient estimates obtained from the IV approach as Local Average Treatment Effects (LATEs) if there are heterogeneous treatment effects. It requires that any patient who is (not) batched by a sequencer (batcher) would also (not) be batched by a batcher (sequencer) physician. The literature leveraging the judges design typically performs two informal tests for its implications. The first one provides that the first stage should be weakly positive for all subsamples (Dobbie et al. (2018)). The second implication asserts that the instrument constructed by leaving out a particular subsample has predictive power over that same left-out subsample (Bhuller et al. (2020)). Appendix Table 8 presents both of these tests in the two columns for various subsamples of interest.

4. Results

4.1. Reduced-Form Results

In this section, we explore the causal influence of physician batch tendency on patient outcomes and resource utilization in the emergency department. We posit that while batch tendency directly influences the practice of batch ordering tests, both batch tendency and batch ordering are concurrently influenced by a physician's testing inclination. Given that testing inclination directly affects primary outcomes, we include it as a control variable in our regression models to mitigate its confounding effects.

To quantify a physician's testing inclination, we employ a similar approach to that used in measuring physician batch tendency. Specifically, we calculate the physician's residualized leave-out average test rate, which serves as a proxy for their propensity to order tests. It is important to note the strong positive correlation ($r = 0.79$) between batch tendency and testing inclination. This substantial correlation suggests that physicians with a higher propensity to test may also exhibit a higher tendency to batch orders, potentially as a consequence of their testing strategies. Neglecting to account for testing inclination could lead to overestimated effects of batch tendency due to omitted variable bias.

Reduced-form regression results in Table **??** reveal the effect the average total effect of batch tendency on resource utilization and LOS. However, it is important to note that the estimated total effect of batch tendency accounts for both direct and indirect (mediated) effects of batch tendency. We hypothesize that while batch ordering may streamline the testing process, resulting in a quicker completion of a given number of tests, it simultaneously appears to lead to an increase in the total number of tests ordered. The increased testing volume, in turn, is associated with an extended LOS.

Table 4 Reduced-Form Results: Length of Stay and Number of Tests

	ln(LOS)	Number of Tests
	(1)	(2)
Batch Tendency	-0.109	0.149^{***}
	(0.464)	(0.048)
Testing Inclination	0.605^{**}	0.952^{***}
	(0.240)	(0.023)
Seasonality and shift fixed effects?	Yes	Yes
Chief Complaint?	Yes	Yes
N	41,929	41,929
R^2	0.215	0.265
Adjusted R^2	0.213	0.263
Residual Std. Error (df = 41800)	0.490	0.842

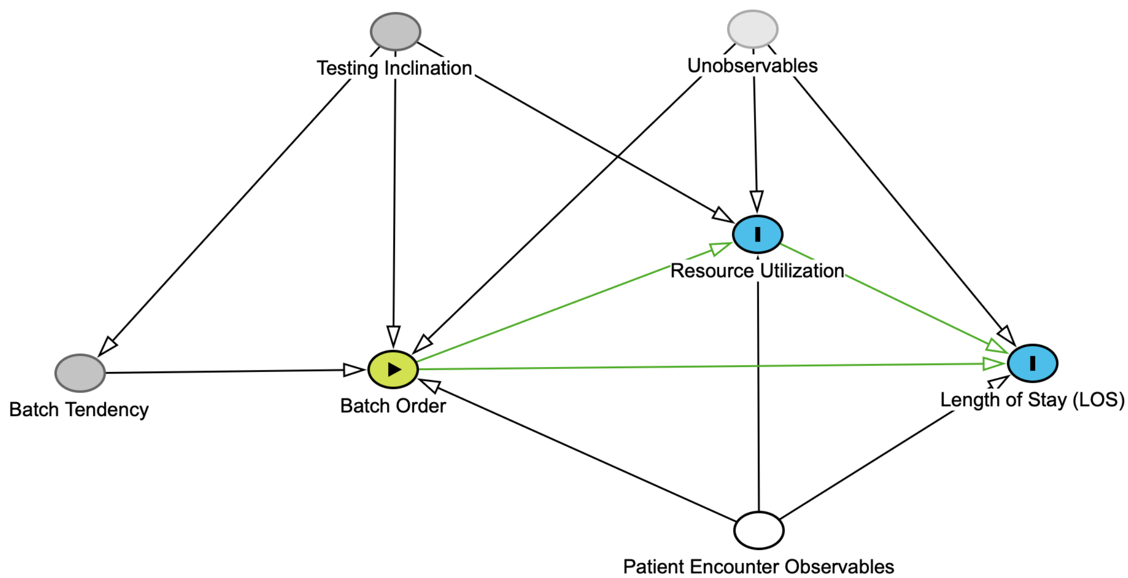
Estimates of the reduced form are for the baseline sample described in the text. Seasonality shift fixed effects include Year-Month and Hospital-Day of week-Hour of day fixed effects. Chief complaint comes from the cleaned complaint that the patient came in with at the initial encounter. Robust standard errors are clustered at the physician level.

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

We introduce a Directed Acyclic Graph (DAG) to illustrate the interconnected pathways between batch tendency, batch ordering, and primary outcomes. The DAG in Figure 2 shows that batch tendency directly influences batch ordering, which in turn affects the number of tests ordered. The number of tests ordered, in turn, influences the length of stay. The DAG also illustrates the effect of batch tendency on the length of stay is mediated by the number of tests ordered.

To determine whether or not the effect of batch tendency on LOS is mediated by the number of tests ordered, we estimate a mediation analysis using the approach outlined by **?**. Our analysis, scrutinizes the pathways through which physician batch tendency influences patient length of stay (LOS) in the emergency department. The analysis delineates both direct effects, which capture the influence of batch tendency on LOS without intermediaries, and indirect effects, which operate through the mediator of test ordering volume. Our findings underscore a significant Average Direct Effect (ADE) of batch tendency on LOS, estimated at -0.15 (p = 0.046).

Figure 2 Directed Acyclic Graph



4.2. Placebo Check

In this section we investigate whether the reduced-form effects observed in Section 4.1 are due to differences in batch rates across providers or due to other provider differences correlated with batch tendency. We start by studying reduced-form effects among patients with complaints that are never batched, as a "placebo/falsification check." By way of example, consider a patient who arrives at the ED with a urinary tact infection—a condition for which patients rarely undergo imaging testing. For such patients, we should expect to see no impact of batch tendency only if high batching and low batching physicians do not systematically differ in other dimensions of care relevant to patient outcomes. Conversely, if we do find a reduced-form effect for these patients, then high batch tendency physicians must systematically differ from low batch tendency physicians in other dimensions of care, beyond batching.

To that end, we restrict attention to ED visits for complaints where batching occurs no more than 10 percent of the time (recall that our baseline sample only includes complaints with a >10 percent batching rate). We estimate a reduced-form regression of each main outcome on physician batch tendency for the subsample, following equation (3). The results of this exercise are displayed in Table 5. They show that in contrast to results for our main sample, the association between physician tendency to batch and a given outcome is statistically indistinguishable from zero and much smaller in magnitude for the samples of patients who visit the ED with health conditions that are rarely batched.

Table 5 Placebo Exercise: Reduced-Form Results for Rarely Batched Samples

	ln(LOS) (1)	Number of Tests (2)	72 Hour Return (3)
Batch Tendency	−0.465 (0.645)	−0.699 (0.427)	−0.011 (0.095)
Seasonality and shift fixed effects?	Yes	Yes	Yes
Chief Complaint?	Yes	Yes	Yes
Testing inclination?	Yes	Yes	Yes
N	2,361	2,361	2,361

This table reports the estimated coefficients of a reduced-form regression of our main outcomes on physician batching tendency for samples based on the patient's chief complaint field. Estimates come from regression on conditions that are batched less than 10 percent of the time. See text for residualization fixed effects and baseline controls. Robust standard errors are clustered at the physician level.
$***p < 0.001$, $**p < 0.01$, $*p < 0.05$.

4.3. Instrumental Variables Results

In this section we examine the causal effects of batch ordering in the ED. Mirroring our presentation of the reduced-form results, we turn to presenting effects in Table **??** on our main outcomes of ED LOS and number of tests ordered. We first start by estimating the effects of batch ordering on the number of tests ordered using the standard 2SLS approach. Next we estimate the direct and indirect effects of batch ordering on LOS using a causal mediation analysis.

ED length of stay (LOS).– Our causal mediation analysis reveals that the Average Direct Effect (ADE) of test batching on the natural log of LOS is -0.13607, indicating that, when not considering the mediating effect of the number of tests, batching leads to a decrease in the expected LOS. Specifically, this translates to an approximate 12.7 percent reduction in LOS when tests are batched, all else being equal. This result, complemented by our 2SLS estimate of -0.086, suggests that the direct effect of batch ordering on LOS is negative. Our results also suggest a "washing out" effect, where the effeciency gains from batching are offset by the increased number of tests ordered. This is consistent with our finding that batching increases the number of tests ordered.

Resource Utilization.– The impact of batching on resource utilization is pronounced. We observe a 0.135 increase in the number of tests ordered due to batching, corresponding to approximately 13.5 excess tests per 100 batch orders, where excess means tests that would not have been ordered had the physician not batched. This result is consistent with the notion that batching leads to increased resource utilization in the ED.

The instrumental variable analysis presented here underscores the complexities underlying batch ordering practices in the ED. While batching is often seen by physicians as a method to enhance

Table 6 Combined 2SLS and Causal Mediation Analysis Results

	2SLS Results	
	ln(LOS)	Number of Tests
Batch	-0.086	0.135***
	(0.419)	(0.046)
Seasonality and Shift Fixed Effects	Yes	Yes
Chief Complaint and Severity	Yes	Yes
Observations	41,901	41,901
R^2	0.183	0.310

	Causal Mediation Analysis	
Effect	Estimate	95% CI
ACME	0.036	[-0.009, 0.08]
ADE	-0.136	[-0.221, -0.04]
Total Effect	-0.100	[-0.195, 0.00]
Proportion Mediated	-0.333	[-2.831, 1.21]

The 2SLS results are based on robust standard errors clustered at the physician level, with controls for seasonality, shift fixed effects, chief complaint, and severity. The causal mediation analysis examines whether the number of tests mediates the effect of batching on length of stay, using a quasi-Bayesian confidence interval approach with 1000 simulations.

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

efficiency, our findings suggest that it may lead to increased resource utilization and that the corresponding gains in patient throughput are washed out. These insights point to the need for a more nuanced approach to test ordering in emergency care, one that carefully weighs the benefits of comprehensive assessment against the risks and costs of potential over-testing.

4.4. Robustness

The 2SLS results presented in the previous sections are robust to several alternative specifications probing the key batching definition. Our primary definition operationalizes batching as the ordering of multiple diagnostic tests within a 5-minute window. This approach is grounded in standard emergency medicine practices and offers a clear distinction between batching and non-batching behaviors. To ensure the robustness of our findings, we performed several sensitivity analyses.

(1) Variation in the Time Window for Batching: Firstly, we varied the time window for what constitutes a batch. The original 5-minute window was extended and contracted to understand its impact on the study's outcomes. Specifically, we considered scenarios where a batch is defined as two or more tests ordered within windows of 1 minute to 10 minutes. This range allowed us to capture a broader spectrum of batching behaviors and test the sensitivity of our results to different batching definitions.

(2) Refinement of Batching Definition: Secondly, we refined our batching definition based on the sequence and timing of test orders. In the initial analysis, if two tests were ordered upfront and another test ordered later, it was counted as a batch. We modified this definition to consider a set of tests as a batch only if all tests ordered during a patient encounter came from that initial batch order. This adjustment ensures that our batching definition more accurately reflects a comprehensive diagnostic effort at the outset of patient care, rather than incremental decision-making.

The key finding from these robustness checks is the consistent impact of batching on key outcomes across all variations. As seen in Table 9 and 10, altering the time window for batching, whether narrowing it to as little as 1 minute or expanding it to 10 minutes, did not qualitatively change the results. Similarly, refining the definition of batching to consider only those tests ordered in the initial batch also had negligible impact on the study's main findings. These results lend further credibility to our initial findings, demonstrating that our conclusions about the impact of batching on patient outcomes in the ED are not sensitive to the specific operational definition of batching.

5. Conclusion

Our results highlight that changes in resource utilization and costs can arise as a consequence of small variations in physician test ordering strategies in a single medical encounter, at the emergency department. Our instrumental variable approach suggests that the causal effects of batching are substantial: 13.5 excess tests arise per 100 batched orders. This is particularly pronounced in high-volume complaint areas, where batching leads to a significant increase in the number of tests ordered. However, we find no significant impact of batching on ED length of stay. Causal mediation analysis reveals that the direct effect of batching on length of stay is negative, but this effect is offset by the increased number of tests ordered. This suggests that the efficiency gains from batching are washed out by the increased resource utilization.

Our findings have important implications for emergency care practice. While batching is often seen as a method to enhance efficiency, our results suggest that it may lead to increased resource utilization and that the corresponding gains in patient throughput are washed out. These insights point to the need for a more nuanced approach to test ordering in emergency care, one that carefully weighs the benefits of comprehensive assessment against the risks and costs of potential over-testing. Future research should explore the impact of batching on other outcomes, such as patient satisfaction and clinical outcomes, to provide a more comprehensive understanding of the implications of batching in the emergency department.

Appendix. General appendix

Table 7: Chief Complaints

Complaint Area	**Complaints**
Abdominal Complaints	Abdominal Cramping, Abdominal Distention, Dyspepsia, Abdominal Pain, Ascites, Hernia, Abdominal Aortic Aneurysm, Abdominal Injury, Pancreatitis, Umbilical Hernia
Abnormal Test Results	Abnormal Lab, Abnormal Potassium, Abnormal Calcium, ECG Changes, Abnormal ECG, Abnormal Test Result, Blood Infection, Acute Renal Failure, Hypocalcemia, Chronic Renal Failure, Pulmonary Embolism, Abnormal X-ray, Hypoglycemic Unawareness, Elevated Blood Pressure, Abnormal Sodium, Hyperglycemia, Hyponatremia, Platelet Disorders, Anemia, Hypoglycemia, Hypertension, Hypotension, Abnormal Chest Imaging, Abnormal Oximetry, Abnormal Stress Test, Blood Sugar Problem, Hypocalcemia, Hyponatremia
Allergic Reaction	Allergic Reaction, Anaphylaxis
Back or Flank Pain	Back Pain, Back Problem, Flank Pain, Sciatica, Back Injury, Disc Disorder
Breast Complaints	Breast Mass, Breast Pain, Breast Problem, Breast Discharge, Breast Cancer, Breast Discharge, Breast Inflammation
Cardiac Arrhythmias	Atrial Fibrillation, Atrial Flutter, Cardiac Valve Problem, Bradycardia, Irregular Heart Beat, Palpitations, POTS, Ventricular Tachycardia, Rapid Heart Rate, Heart Problem, Cardiac Arrest, Congestive Heart Failure, Circulatory Problem, Transient Ischemic Attack, Ventricular Tachycardia
Chest Pain	Chest Injury, Chest Pain, Chest Wall Pain, Angina, Collarbone Injury, Rib Injury, Heart Pain
Dizziness / Lightheadedness / Syncope	Dizziness, Near Syncope, Syncope, Vertigo, Spells, Hypotension, Paroxysmal Positional Vertigo, Paroxysmal Positional Vertig
Ear Complaints	Cerumen Impaction, Ear Drainage, Ear Fullness, Ear Laceration, Ear Problem, Earache, Hearing Problem, Tinnitus, Ear Injury, Hearing Loss, Nasal Trauma
Epistaxis	Epistaxis, Epistaxis (Nose Bleed), Nose Problem
Exposures, Bites, and Envenomations	Animal Bite, Body Fluid Exposure, Chemical Exposure, Poisoning, Exposure to STD, Insect Bite, Smoke Inhalation, Radiation, Snake Bite, Toxic Inhalation
Extremity Complaints	Ankle Injury, Ankle Pain, Arm Injury, Arm Pain, Cold Extremity, Arm Swelling, Arthritis, Elbow Injury, Elbow Pain, Pseudogout, Extremity Pain, Extremity Weakness, Finger Injury, Hip Injury, Extremity Weakness, Finger Injury, Finger Pain, Dislocation, Foot Infection, Foot Injury, Foot Numbness, Foot Pain, Foot Swelling, Foot Ulcer, Foot Wound Check, Hand Injury, Hand Pain
Eye Complaints	Blurred Vision, Decreased Visual Acuity, Diplopia, Detached Retina, Eye Drainage, Eye Exposure, Eye Pain, Eye Problem, Eye Swelling, Eye Trauma, Foreign Body Eye, Flashes / Light, Loss of Vision, Red Eye, Visual Field Change, Eyelid Problem, Itchy Eye, Eye Exam, Burning Eyes, Eye Twitching, Eyelid/brow Lift Evaluation, Strabismus, Glaucoma, Spots / Floaters
Falls, Motor Vehicle Crashes, Assaults, and Trauma	Assault Victim, Concussion, Facial Injury, Fall, Nasal Trauma, Head Injury, Head Laceration, Motor Vehicle Crash, Puncture Wound, Sexual Assault, Trauma, Domestic Violence, Gun Shot Wound, Work Related Injury, Motorcycle Crash, Injury, Bicycle Accident, Near Drowning, Lip Laceration
Fatigue and Weakness	Difficulty Walking, Fatigue, Gait Problem, Weakness-Generalized, Chronic Fatique, Weakness- Generalized
Fevers, Sweats or Chills	Chills, Diaphoresis, Fever, Night Sweats, Diaphoretic, Diapohresis, Hoarseness, Laryngitis
Foreign Body	Food Bolus, Foreign Body, Foreign Body in Ear, Foreign Body in Skin, Foreign Body in Vagina, Swallowed Foreign Body, Foreign Body in Nose, Foreign Body, FB eye, Foreign Body in Rectum

Gastrointestinal Issues	Anal Fissure, Black or Bloody Stool, Constipation, GERD, Anal Fistula, Diarrhea, Dysphagia, Fecal Impaction, Fistula Follow Up, GIbleeding, GI Problem, Hemorrhoids, Morning Sickness, Nausea, Ostomy Care, Rectal Bleeding, Rectal Pain, Vomiting, Vomiting Blood, Vomiting During Pregnancy, GI Bleeding, Fecal Incontinence, Bloated, Hematochezia, Urine Leakage, Heartburn, Rectal Discharge, Urolithiasis, Ulcerative Colitis, Irritable Bowel Syndrome, Rectal Prolapse, Fistula Evaluation, Rectal Problems, Perianal Abscess, Fisula Evaluation, Stoma Dysfunction
Genital Complaints	Groin Burn, Groin Pain, Groin Swelling, Inguinal Hernia, Menstrual Problem, Pelvic Pain, Penis Pain, Priapism, Testicle Pain, Menorrhagia, Vaginal Bleed, Vaginal Bleeding, Vaginal Itching, Bartholin's Cyst, Genital Warts, Groin Injury, Vaginal Bleeding-Pregnant, Vag Bleed Pregnant, Female Genital Issue, Penis Injury, Vaginal Discharge, Vaginal Pain, Erectile Dysfunction, Vaginal Prolapse, Urethral Stricture, Penile Discharge, Menorrhagia, Gynecologic Exam, Menstrual Problem, Vaginitis/Bacterial Vaginosis, Ovarian Cyst, Vaginitis / Bacterial Vaginosi
Medical Device or Treatment Issue	Cast Problem, Device Check, Dressing Change, Feeding Tube, AICD Problem, Insulin Pump Visit, Gastrostomy Tube Change, Medication Reaction, Shunt, Appliance Removal, Tube Problem, Urinary Catheter Change, Vascular Access Problem, Enteral Nutrition Evaluation, Device Malfunction, Pacemaker Problem, Remova / Exchange Catheter, Drain Removal, Outpatient Infusion, Treatment, Heart Assist Device, Stoma Dysfunction, Tracheostomy Tube Change, Ureteral Stent Exchange
Medication Request	Immunizations, Infusion / Injection Administration, IV Medication, Infusion/ Injection Administ, Med Refill, Medication Visit, Pain Management, Blood Product Administration, Labs Only, Tetanus (Td & Tdap), Wound Care
Neurological Issue	Altered Mental Status, Cognitive Concerns, Facial Droop, Pre Syncope, Focal Weakness, Headache, Memory Loss, Migraine, Dementia, Dysphasia, Neuro Problem, Numbness, Paralysis, Seizures, Slurred Speech, Spasms, Stroke Like Symptoms, Tingling, Tremors, Trigeminal Neuralgia, Unable to Speak, Seizure Disorder, Insomnia, Parkinson's Disease, Loss of Consciousness, Neuropathy, Ataxia, Unable to speak, Peripheral Neuropathy, Stroke, Cerebrovascular Accident, Speech Problem, Acute Neurological Problem, Flashes, Light, Unresponsive, Multiple Sclerosis, Parkinson's Disease, Febrile Seizure, Paresthesia, Peripheral Neuropathy, Hydrocephalus, Spasticity, Neuroendocrine Tumor
Other	Dehydration, Fisula Evaluation, Follow-Up, Illness, Letter for School/Work, Aneurysm, Lung Eval, Error, Mass, Oral Swelling, Other, Advice Only, Deformity, Electric Shock, Personal Problem, Shaking, Swelling, Swollen Glands, Adenopathy, Adrenal Problem, Thrombophilia, Weight Gain, Weight Loss, Hiccups, , Chemo Related Symptoms, Hot Flashes, Follow-up, Non Healing Wound, (Other), Mouth Injury, Xerostomia, Prostate Check, Suture / Staple Removal, Wellness, Voice Changes, Vital Sign Check, Coagulation Disorder, Cold Exposure, Consult, Dental Problem, Tetanus (Td & Tdap), Infusion/ Injection Administ, Tracheostomy Tube Change, Medical Information, Neutropenic Fever, Infection, Leukemia, Heat Exposure, Poor Appetite, Gingivitis, Pre-op Exam, gingivitis, Loss of appetite, Failure To Thrive, Referral, Lymphoma, Hot Flashes, Neutropenia, Radiation, Ingestion, TB Test, Fussy, Lupus, Toxic Inhalation, Lung Screening, Leakage/Loss of Fluid, Liver Eval, Hepatic Cancer, Lung Mass, Venous Thromboembolic Disease, Insulin Pump Visit, Preventive Visit, Avulsion, Peripheral Edema, Hypoglycemic Unawareness, Immobility, Giant Cell Arteritis, Polydipsia, Platelet Disorders, Post-procedure, Lung Follow-up, Poisoning, Injections, POTS, Insulin Reaction, Liver Transplant, Labs Only

Other Pain	Dental Pain, Facial Pain, Generalized Body Aches, Myalgia, Dental Injury, Jaw Pain, Muscle Pain, Neck Pain, Pain, Sickle Cell Pain Crisis, Paresthesia, Torticollis, Chronic Pain, Cancer Pain, Incisional Pain, Bone Pain, Tailbone Pain, Gout, Muscle pain/Weakness, Pseudogout
Post-Op Issue	Post-Op, Post-Procedure, Post-Op Problem, Post-op, Post-Op Issue, Wound Dehiscence, Post-op Problems, Post-op Problem
Psychiatric Complaints	Anxiety, Auditory Hallucinations, Depression, Panic Attack, Homicidal, PTSD (Post-Traumatic Stress, Delusional, Fussy, Paranoia, Suicide Attempt, Hallucinations, Manic Behavior, Eating Disorder, Suicidal, Agitation, Psychiatric Evaluation, Aggressive Behavior, Mental Health Problem, Inappropriate Words
Shortness of Breath	Airway Obstruction, Aspiration, Pain With Breathing, Near Drowning, Respiratory Distress, Shortness of Breath, Wheezing, Increased Work Of Breathing, Difficulty Breathing, Choking, Oxygen Dependence, Hyperventilating, Orthopnea
Skin Complaints	Abrasion, Abscess, Bleeding/Bruising, Blister, Angioedema, Lip Laceration, Burn, Cellulitis, Cyst, Drainage from Incision, Disturb of Skin Sens, Edema, Extremity Laceration, Facial Burn, Cyanosis, Impetigo, Facial Laceration, Facial Swelling, Finger Laceration, Leg Rash, Herpes Zoster, Hives, Itching, Jaundice, Diabetic Ulcer, Diabetic Wound, Laceration, Mouth Lesions, Non-Healing Wound, Rash, Recurrent Skin Infections, Skin Problem, Sore, Scabies, Suture \Staple Removal, Wound Check, Wound Infection, Lesion, Skin Check, Minor Skin Infection, Skin Ulcer, Skin Discoloration, Sunburn, Head Lice, Scabies, Fungal Infection, Leg Rash, Impetigo
Substance Abuse Issues	Alcohol Intoxication, Alcohol Problem, Withdrawal, Drug Overdose, Drug / Alcohol Dependency, Addiction Problem, Addiction Assessment, Delirium Tremens (DTS)
Upper Respiratory Symptoms	Congestion, Cough, Coughing Up Blood, Flu Symptoms, Enlarged Tonsils, Peritonsillar Abscess, Nasal Congestion, Sinus Symptoms, Sinusitis, Sore Throat, Hoarseness, Throat Problem, Upper Respiratory Infection, Influenza, Laryngitis, Respiratory Arrest, Pneumonia, Pleural Effusion, Asthma, Croup, URI, Peritonsillar Abscess
Pregnancy Related	Pregnancy Problem, Miscarriage, Contractions, Ectopic Pregnancy, Laboring, Possible Pregnancy, Pregnancy Related
Renal	Av Fistula, Kidney Transplant, Elevated Serum Creatinine, End-Stage Liver Disease, Hemodialysis Access, Nephritis, Ureteral Stent Exchange
Urinary Complaints	Bladder Problem, Blood in Urine, Cystitis, Difficulty Urinating, Dysuria, Gross Hematuria, Painful Urination, Urinary Frequency, Urinary Symptom, Urinary Incontinence, Urinary Problem, Urinary Retention, Slowing Urinary Stream, Urinary Tract Infection, Urinary Urgency, Voiding Dysfunction, Hesitancy Urinary

Table 8 Testing the Monotonicity Assumption

Sub-sample margin	Baseline Tendency	Reverse-Sample Tendency
Extremity Complaints	0.864*** (0.128)	0.893*** (0.068)
Upper Respiratory Symptoms	1.235*** (0.247)	0.953*** (0.038)
Abnormal Test Results	0.980*** (0.219)	0.766*** (0.111)
Falls, Motor Vehicle Crashes, Assaults, and Trauma	1.061*** (0.167)	0.784*** (0.103)
Shortness of Breath	1.204*** (0.232)	0.944*** (0.047)
Gastrointestinal Issues	1.049*** (0.140)	0.929*** (0.052)
Chest Pain	1.073*** (0.223)	0.943*** (0.048)
Dizziness/Lightheadedness/Syncope	1.899*** (0.171)	0.930*** (0.042)
Back or Flank Pain	0.657*** (0.180)	0.718*** (0.106)
Neurological Issue	0.970*** (0.191)	0.835*** (0.067)
Fatigue and Weakness	1.698*** (0.230)	0.938*** (0.058)
Cardiac Arrhythmias	2.571*** (0.243)	1.107*** (0.059)
Abdominal Complaints	0.903*** (0.275)	0.971*** (0.019)
Skin Complaints	0.448*** (0.098)	0.584*** (0.108)
Fevers, Sweats or Chills	1.033*** (0.305)	0.888*** (0.084)

Column 1 displays the first stage coefficient of batched on the baseline physician batch tendency instrument for the corresponding sub-sample. Column 2 constructs a new physician batch tendency instrument using all emergency visits, excluding the corresponding sub-sample ("reverse-sample"), and displays the coefficient of the first stage regression back on that sub-sample. Robust standard errors are clustered at the physician level.

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

Table 9 2SLS Results: Robustness to Batching Definition (1-minute and 10-minute windows)

	1-minute window		
	Log LOS	Number of Tests	72 Hour Return
	(1)	(2)	(3)
Testing Inclination	0.592***	0.975***	−0.025***
	(0.203)	(0.018)	(0.009)
Batch	−0.101	0.159***	0.021
	(0.496)	(0.054)	(0.023)
N	41,929	41,929	41,929
R^2	0.209	0.278	0.009
Adjusted R^2	0.204	0.274	0.003
Residual Std. Error	0.492	0.836	0.184
	10-minute window		
	Log LOS	Number of Tests	72 Hour Return
	(1)	(2)	(3)
Testing Inclination	0.650***	0.957***	−0.027**
	(0.241)	(0.022)	(0.011)
Batch	−0.205	0.141***	0.017
	(0.438)	(0.045)	(0.018)
N	41,929	41,929	41,929
R^2	0.146	0.315	0.009
Adjusted R^2	0.141	0.311	0.003
Residual Std. Error	0.512	0.814	0.184

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

Table 10 2SLS Results: Robustness to Batching Definition (all tests at once)

	Log LOS	Number of Tests	72 Hour Return
	(1)	(2)	(3)
Batch	0.592***	0.975***	−0.025***
	(0.203)	(0.018)	(0.009)
‘any.batch(fit)‘	−0.101	0.159***	0.021
	(0.496)	(0.054)	(0.023)
N	41,929	41,929	41,929
R^2	0.209	0.278	0.009
Adjusted R^2	0.204	0.274	0.003
Residual Std. Error (df = 41667)	0.492	0.836	0.184

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 11 2SLS Results for the "Batch" Variable Across Different General Complaints: Outcome of Number of Tests

Complaint	Batching Coefficient	Standard Error
Musculoskeletal/Extremity	0.476**	(0.235)
Respiratory-Related	0.244	(0.209)
General/Other Symptoms	0.179	(0.207)
Gastrointestinal/Abdominal	-0.092	(0.128)
Cardiac/Chest-Related	0.270	(0.264)
Neurological/Syncope	-0.080	(0.493)

Acknowledgments

The authors would like to thank ..

References

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

Bernstein S, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, McCarthy M, McConnell J, Pines J, Rathlev N, et al. (2009) The effect of emergency department crowding on clinically oriented outcomes. *Academic Emergency Medicine* 16(1):1–10.

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

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`.

Koch C, Roberts K, Petruccelli C, Morgan D (2018) The frequency of unnecessary testing in hospitalized patients. *Am J Med* 131(5):500–503, URL `http://dx.doi.org/10.1016/j.amjmed.2017.11.025`, pMID: 29224739; PMCID: PMC8628817.

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.

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