Appendix

A. ED Arrival Patterns and Capacity

Figure A1: Arrival Rate by Hour

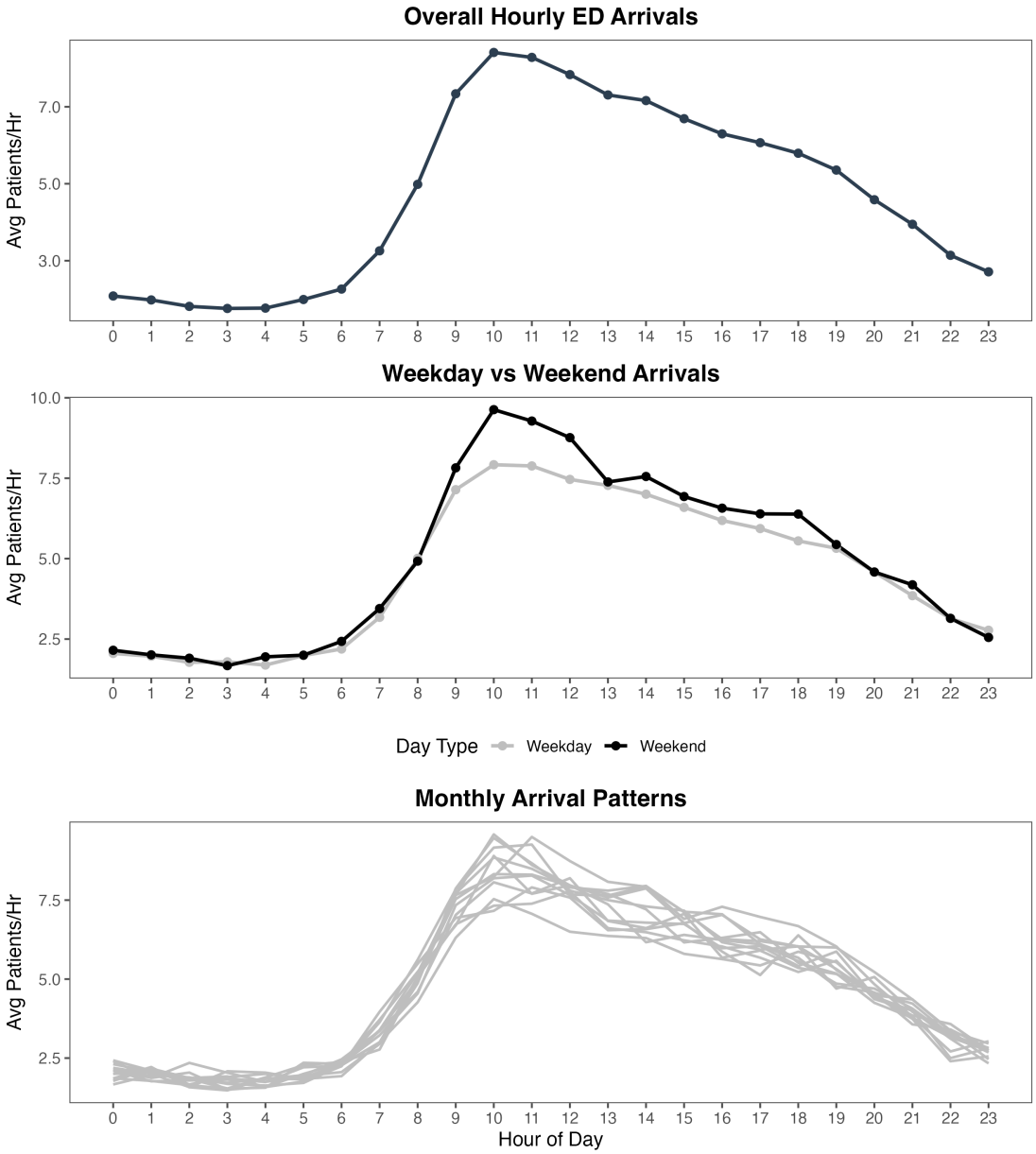
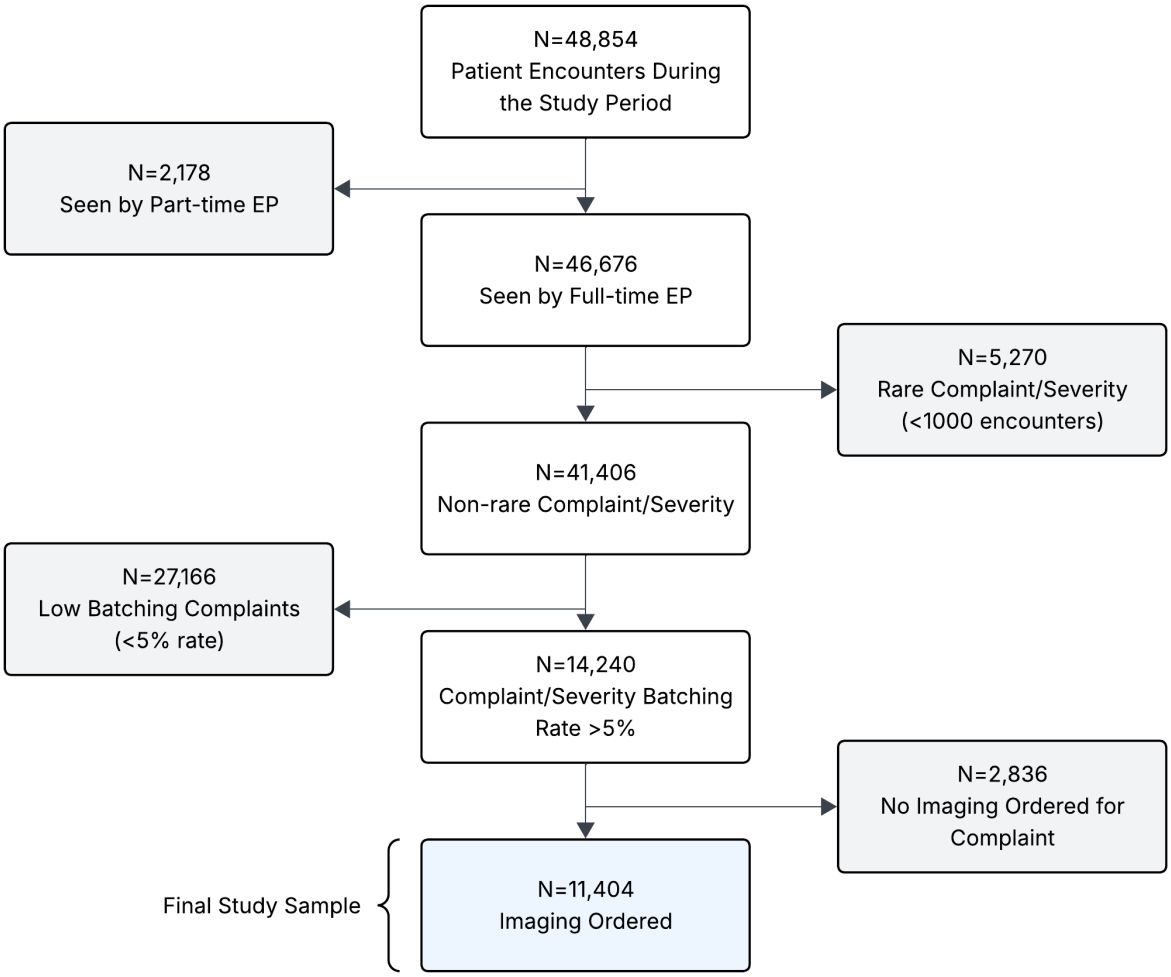


Table A1: Internal Guidelines for ED Overcapacity and Saturation Activation

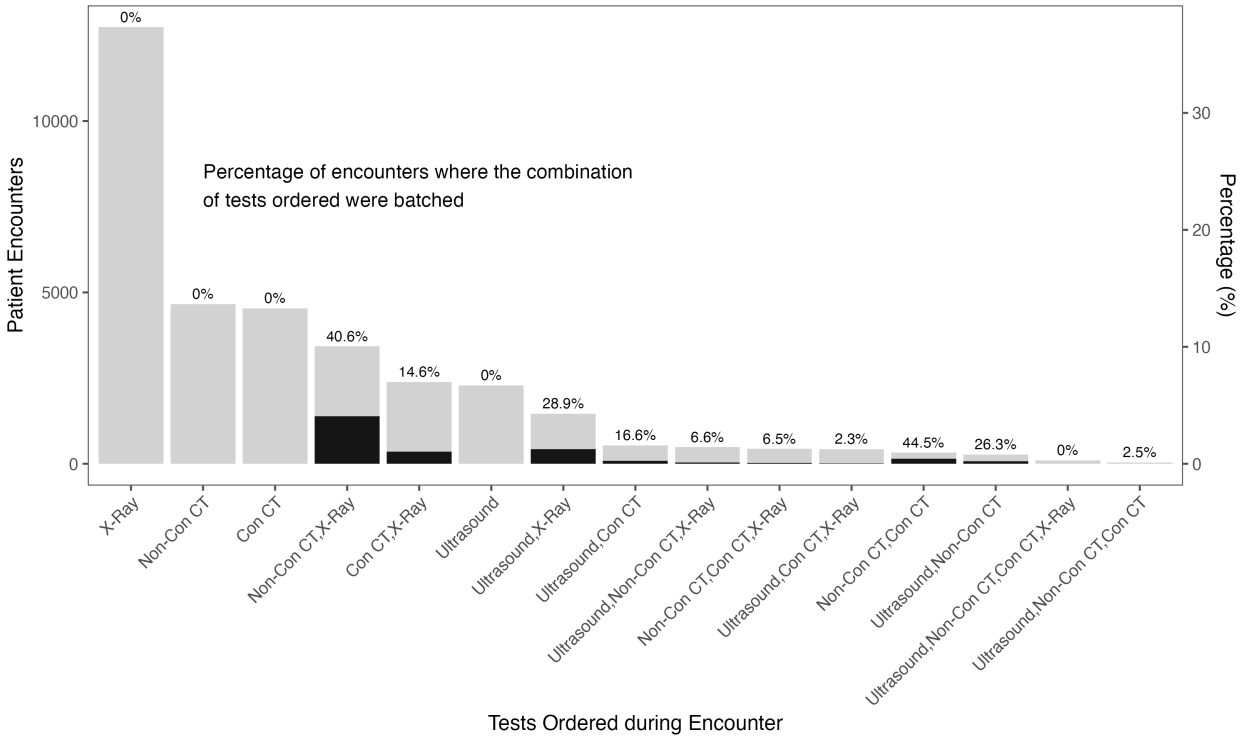
Capacity Level	**Conditions**	**Actions**
Normal Operations	• Waiting room time < 20 minutes • ED census below staffed bed capacity	• No RMA or waiting-room evaluations unless treatment room unnecessary • No waiting-room laboratory testing • Nurse-initiated protocols entered 30 minutes after arrival
Minor Overcapacity	• Waiting room time 21–90 minutes • $\geq$10 patients waiting • ED census exceeds beds by $\leq$10 patients • Team Lead discretion	• All normal-operations actions • Waiting-room lab initiation • RMA activation with dedicated staff • Expanded zone placement • Expedited consulting communication
Major Overcapacity	• Waiting room time > 90 minutes • $>$20 patients waiting • $>$40 arrivals in 2 hours • ED census exceeds beds by $\geq$20 patients • Team Lead discretion	• All minor-overcapacity actions • Consider diversion • Second RMA activation if feasible • Hospital-wide throughput escalation • Activation of observation APPs and on-call ED physician

Figure A2: CONSORT Flow Diagram for Sample Selection



Notes: This figure displays the sample selection process for the Mayo Clinic ED data. Starting with 48,854 patient encounters during the study period (October 2018 - December 2019), we apply sequential exclusion criteria to arrive at our analytical sample of 11,404 encounters. Exclusions are necessary to ensure sufficient variation in batching behavior for instrumental variable identification. Rare complaints are those with fewer than 1,000 total encounters. Low-batching complaints are those where batching occurs in less than 5% of encounters.

Figure A3: Distribution of imaging test combinations and batching rates



Notes:

B. Complier, Always-Taker, and Never-Taker Classification

This section describes the method used to calculate the share of compliers, always-takers, and never-takers in the context of batch ordering in the emergency department (ED). Our approach follows Dahl et al. [2014], Dobbie et al. [2018], and Eichmeyer and Zhang [2022].

Overview

Compliers are defined as patients who would not have had their imaging tests ordered in a batch if they had been seen by a low-batch-tendency physician ("sequence") but would have had their imaging tests batched if they had been seen by a high-batch-tendency physician ("batcher"):

$$\pi_{\text{complier}} = P(B_i = 1|Z_i = \bar{z}) - P(B_i = 1|Z_i = \underline{z}) = P(B_{\bar{z}i} > B_{\underline{z}i})$$

where B_i represents the batch ordering decision for patient i, Z_i represents the batch tendency of patient i's assigned physician, and $\bar{z}$ and $\underline{z}$ represent the maximum and minimum values of our batch tendency instrument (the highest and lowest batch tendency physicians), respectively.

Always-takers are patients whose imaging tests would be batched regardless of which physician they see. Because of the monotonicity and independence assumptions, the fraction of always-takers is given by the probability of being batched by the most conservative (lowest batch tendency) physician:

$$\pi_{\text{always-taker}} = P(B_i = 1|Z_i = \underline{z}) = P(B_{\bar{z}i} = B_{\underline{z}i} = 1)$$

Finally, never-takers are patients whose imaging tests would never be batched regardless of which physician they see, with the fraction of never-takers given by the probability of not being batched by the most aggressive (highest batch tendency) physician:

$$\pi_{\text{never-taker}} = P(B_i = 0|Z_i = \bar{z}) = P(B_{\bar{z}i} = B_{\underline{z}i} = 0)$$

Number of Compliers

We calculate the shares of patients in each category by examining batch ordering rates for patients assigned to physicians at different points in the batch tendency distribution. Following Dahl et al. [2014], we define the "most aggressive" batch-ordering physicians ($\bar{z}$) as those at the top 1 percentile of batch tendency and the "most conservative' ' batch-ordering physicians ($\underline{z}$) as those at the bottom 1 percentile.

In the first three columns of Table B1, we estimate a local linear regression of $Batched_i$ on our residualized measure of physician batch tendency. Under this more flexible analog to our first-stage equation, we find that approximately 20 percent of our sample are compliers, 73 percent are never-takers, and 7 percent are always-takers.

In the last three columns of Table B1, we estimate our linear specification of the first stage, given by Equation (4). Under this specification, we can recover π_c as $\hat{\alpha}_1(\bar{z} - \underline{z})$, π_a as $\hat{\alpha}_0 + \hat{\alpha}_1\underline{z}$, and π_n as $1 - \hat{\alpha}_0 - \hat{\alpha}_1\bar{z}$, where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are the estimated first-stage coefficients. Under this linear specification, we find that 21 percent of our sample are compliers, 72 percent are never-takers, and 7 percent are always-takers. We also explore the sensitivity of the estimated share of compliers, always-takers, and never-takers to the exact choice of cutoff for the most aggressive and most conservative physicians. As shown in Table B1, our results are robust to the particular model specification and cutoff.

Table B1: Sample Share by Compliance Type

	Local Linear Model			Linear Model		
Batch Tendency Cutoff:	1%	1.5%	2%	1%	1.5%	2%
Compliers	0.20	0.20	0.20	0.21	0.21	0.21
Never-Takers	0.73	0.73	0.73	0.72	0.72	0.72
Always-Takers	0.07	0.07	0.07	0.07	0.07	0.07

Notes: This table presents the estimated share of compliers, never-takers, and always-takers under different model specifications and batch tendency cutoffs. The local linear model estimates a loess regression of batching on batch tendency. The linear model estimates Equation (4) in the main text. Cutoffs define the percentiles used to identify the most aggressive (top percentile) and most conservative (bottom percentile) batch-ordering physicians.

Characteristics of Compliers

We also characterize our population of compliers by observable characteristics, which can be recovered by calculating the fraction of compliers in different subsamples [Abadie and Gardeazabal, 2003, Dahl et al., 2014]. For any binary characteristic X_i, we compute:

$$P(X_i = x|\text{complier}) = \frac{\pi_{c|x} \cdot P(X_i = x)}{\pi_c}$$

where $\pi_{c|x}$ is the complier share estimated within the subsample where $X_i = x$, and π_c is the overall complier share. The ratio $P(X_i = x|\text{complier})/P(X_i = x)$ indicates whether compliers are over-represented (ratio > 1) or under-represented (ratio < 1) among patients with characteristic $X_i = x$.

Table B2 presents the sample distribution, complier distribution, and relative likelihood for different patient subgroups. Several patterns emerge. Compliers are significantly more likely to be female (ratio = 1.21) and to have normal vital signs at presentation (ratio = 1.13). Compliers are less likely to be tachycardic (ratio = 0.57) or to have any abnormal vital sign (ratio = 0.79). Compliers are also more likely to have laboratory tests ordered (ratio = 1.13) but less likely to present without laboratory orders (ratio = 0.49).

These patterns suggest that compliers—patients at the margin of physician discretion—tend to present with less acute clinical pictures. Patients with abnormal vital signs or without laboratory workups may have more obvious clinical trajectories that dictate a particular imaging strategy regardless of physician preference. In contrast, patients with normal vitals and standard laboratory workups represent the diagnostic "gray zone'' where physician practice style most influences testing decisions.

Table B3 presents analogous results by chief complaint category. Compliers are substantially over-represented among patients presenting with Falls, Motor Vehicle Accidents, Assaults, and Trauma (ratio = 1.66) and Neurological Issues (ratio = 1.24). Compliers are under-represented among patients with Abdominal Complaints (ratio = 0.62) and Fevers, Sweats, or Chills (ratio = 0.62). This pattern is consistent with clinical intuition: trauma and neurological presentations often involve diagnostic uncertainty about the extent of injury, creating opportunities for physician discretion in imaging strategy. In contrast, abdominal pain and fever presentations may have more standardized imaging protocols that reduce the influence of individual physician preference.

Together, these complier characteristics help interpret our LATE estimates. Our instrumental variables analysis identifies the effect of batch ordering for the approximately 20 percent of patients whose imaging strategy depends on physician preference rather than clinical necessity. These marginal patients tend to be older, female, with normal vital signs and standard laboratory workups, and are disproportionately likely to present with trauma or neurological complaints. This population represents exactly the "gray zone'' where ED operational interventions could influence practice patterns without constraining clinically necessary care.

Table B2: Characteristics of Marginal Patients (Compliers)

	$P[X = x]$	$P[X = x\|\text{complier}]$	$\frac{P[X=x\|\text{complier}]}{P[X=x]}$
Demographics			
Male	0.477	0.407	0.853
Female	0.523	0.633	1.211
White	0.898	0.918	1.022
Non-White	0.102	0.120	1.181
Age			
Age < 50	0.230	0.204	0.887
Age ≥ 50	0.770	0.834	1.082
Acuity			
High Acuity (ESI 1–2)	0.486	0.478	0.984
Lower Acuity (ESI 3–5)	0.514	0.560	1.090
Vital Signs			
Tachycardic	0.201	0.115	0.569
Any Abnormal Vital	0.280	0.222	0.793
Normal Vitals	0.720	0.817	1.134
Laboratory Testing			
Labs Ordered	0.780	0.877	1.125
No Labs Ordered	0.220	0.108	0.492

Notes: This table presents the sample distribution, complier distribution, and relative likelihood for different patient subgroups. Column 1 reports the unconditional probability of each characteristic in the full sample. Column 2 reports the probability of each characteristic among compliers, estimated following Abadie and Gardeazabal [2003]. Column 3 reports the ratio of the complier probability to the sample probability, indicating whether compliers are over-represented (ratio > 1) or under-represented (ratio < 1) among patients with each characteristic.

Table B3: Complier Characteristics by Chief Complaint

Chief Complaint	$P[X = x]$	$P[X = x\|\text{complier}]$	$\frac{P[X=x\|\text{complier}]}{P[X=x]}$
Neurological Issue	0.208	0.259	1.241
Falls, MVA, Assaults, Trauma	0.171	0.284	1.656
Extremity Complaints	0.228	0.199	0.870
Dizziness/Lightheadedness/Syncope	0.106	0.094	0.882
Fatigue and Weakness	0.103	0.091	0.887
Abdominal Complaints	0.100	0.062	0.623
Fevers, Sweats, or Chills	0.083	0.052	0.623

Notes: This table presents the sample distribution, complier distribution, and relative likelihood for each chief complaint category. Compliers are patients whose batching decision depends on the assigned physician's batch tendency. See notes to Table B2 for estimation details.

C. Mediation Analysis: SEM Results

Table C1: Mediation Analysis of the Effect of Batching on ED Length of Stay (LOS) via Imaging and Admission Decisions

	Estimate	Standard Error	p-value
Number of Imaging Tests (Residualized)			
Batching (b_1)	1.434	0.106	<0.001
Tachycardic	0.077	0.015	<0.001
Tachypneic	0.111	0.020	<0.001
Febrile	0.051	0.028	0.062
Hypotensive	0.111	0.038	0.003
Arrival Age (scaled)	0.002	0.000	<0.001
Admission Decisions (Residualized)			
Batching (b_2)	0.327	0.069	<0.001
Number of Imaging Tests (b_3)	0.162	0.006	<0.001
Tachycardic	0.084	0.010	<0.001
Tachypneic	0.056	0.013	<0.001
Febrile	0.142	0.018	<0.001
Hypotensive	0.210	0.024	<0.001
Arrival Age (scaled)	0.002	0.000	<0.001
Length of Stay (Residualized)			
Number of Imaging Tests (c_1)	0.144	0.006	<0.001
Admission Decisions (c_2)	0.089	0.010	<0.001
Batching Direct Effect (c')	0.031	0.072	0.662
Tachycardic	0.041	0.010	<0.001
Tachypneic	-0.005	0.013	0.721
Febrile	-0.020	0.018	0.283
Hypotensive	-0.025	0.025	0.326
Arrival Age (scaled)	0.002	0.000	<0.001
Indirect and Total Effects			
Indirect via Imaging Tests ($b_1 \times c_1$)	0.207	0.018	<0.001
Indirect via Admission Decisions ($b_2 \times c_2 + b_1 \times b_3 \times c_2$)	0.050	0.008	<0.001
Total Indirect Effect	0.256	0.020	<0.001
Total Effect	0.288	0.073	<0.001

Notes: This table presents the results of a structural equation model (SEM) investigating the relationships between batching, the number of imaging tests, admission decisions, and ED length of stay (LOS). To address potential confounding from fixed effects (e.g., complaint type and time of month), all variables were residualized by regressing them on the fixed effects prior to the SEM. This residualization ensures that the estimates reflect associations net of the fixed effects.

Table C2: Mediation Analysis of the Effect of Batching on Time to Disposition

	Estimate	Standard Error	p-value
Number of Imaging Tests (Residualized)			
Batching (b_1)	1.434	0.106	<0.001
Tachycardic	0.077	0.015	<0.001
Tachypneic	0.111	0.020	<0.001
Febrile	0.051	0.028	0.062
Hypotensive	0.111	0.038	0.003
Arrival Age (scaled)	0.002	0.000	<0.001
Time to Disposition (Residualized)			
Number of Imaging Tests (c_1)	0.060	0.007	<0.001
Batching Direct Effect (c')	-0.100	0.081	0.216
Tachycardic	0.006	0.011	0.617
Tachypneic	-0.018	0.015	0.236
Febrile	-0.098	0.021	<0.001
Hypotensive	-0.090	0.028	0.002
Arrival Age (scaled)	-0.001	0.000	0.042
Indirect and Total Effects			
Indirect via Imaging Tests ($b_1 \times c_1$)	0.085	0.012	<0.001
Total Effect	-0.015	0.081	0.855

Notes: This table presents results from a structural equation model (SEM) analyzing the relationship between batching, the number of imaging tests, and time to disposition. All variables were residualized to account for fixed effects of day of the week, month, and chief complaint by severity. The coefficients for the number of imaging tests reflect its role as a mediator in the pathway from batching to time to disposition. The direct effect of batching is not significant, while the indirect effect via imaging tests is significant and positive, supporting the hypothesis that increased diagnostic intensity prolongs time to disposition. This analysis is exploratory and intended to provide insight into plausible mechanisms, not causal mediation.

D. Robustness Checks

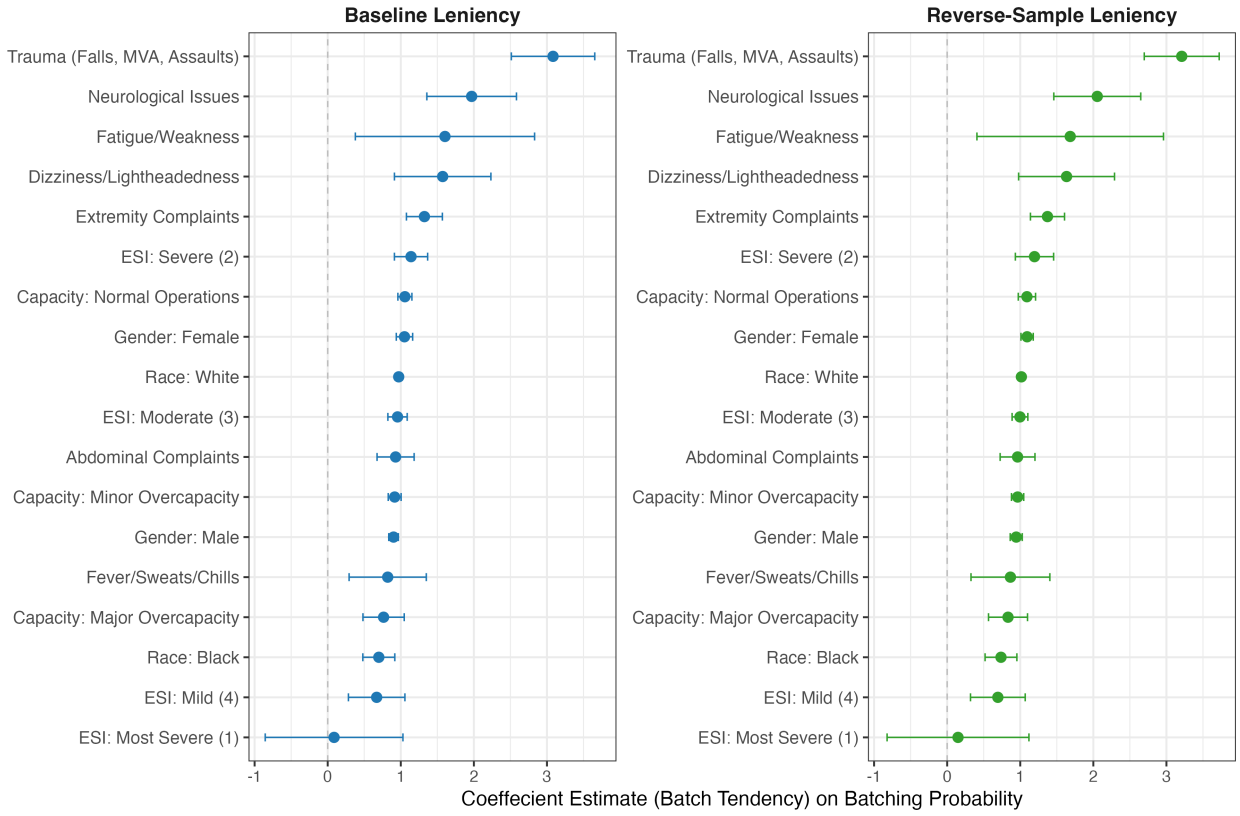


Figure D1: Testing the Monotonicity Assumption

Notes: The left panel displays the first stage coefficient and 95% CI of batching on the baseline physician batch tendency instrument for the corresponding sub-sample. The right panel constructs a new physician batch tendency instrument using all emergency visits, excluding the corresponding sub-sample ("reverse-sample"), and displays the coefficient and 95% CI of the first stage regression back on that sub-sample. Robust standard errors are clustered at the physician level.

Table D1: Placebo Check: Batch Tendency and Patient Outcomes

	Dependent variable		
	Log time to disposition (1)	Log LOS (2)	72hr return with admission (3)
Batch tendency	3.486	3.580	−0.2158
	(1.724)	(1.750)	(0.1069)
Mean dependent variable	4.57	4.59	0.001
	(0.611)	(0.615)	(0.098)
Time FE	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes
Adj R^2	0.159	0.160	0.051
Observations	1,244	1,244	1,244

Notes: This table reports the estimated coefficients of a reduced-form regression of our main outcomes on physician batch tendency for patients with conditions for which batching occurs less than 1% of the time. The dependent variables are log time to disposition, log LOS, and 72-hour return with admission. Robust standard errors are clustered at the physician level.

Placebo Exercise

We investigate whether the reduced-form effects observed in Section 4.1 are due to differences in batch ordering rates across physicians or due to other provider differences correlated with batching tendency. We examine this by studying reduced-form effects among patients with conditions that rarely require multiple imaging tests, as a falsification check. For example, consider a patient who arrives at the ED with an isolated ankle injury—a condition for which multiple imaging tests are rarely ordered. For such patients, we should expect to see no impact of physician batch tendency if "batchers" and "sequencers" do not systematically differ in other dimensions of care relevant to patient outcomes.

We restrict attention to ED visits for conditions which batching occurs less than 1 percent of the time, as a placebo check. We estimate reduced-form regressions of each main outcome on physician batch tendency for this subsample and the results are presented in Table D1. The results show no significant association between physician batch tendency and patient outcomes for this subsample, providing evidence that the reduced-form effects observed in the main analysis are not driven by unobserved physician differences correlated with batching tendency.

This placebo analysis supports our identification strategy by suggesting that batch tendency is not systematically correlated with other physician practice patterns that might affect patient outcomes. However, we recognize that given the multidimensionality of physician behavior, we also present reduced-form results and maintain focus on imaging-related outcomes, for which such concerns are less pronounced.

Table D2: Robustness Check for Main 2SLS Results

	(1)	(2)
Panel A. Primary Outcomes		
Log time to disposition	0.923^{*}	$0.753^{\dagger}$
	(0.358)	(0.368)
Log LOS	0.783^{*}	0.656^{*}
	(0.300)	(0.296)
Number of distinct imaging tests	1.271^{***}	1.307^{***}
	(0.218)	(0.213)
72hr return with admission	-0.005	-0.015
	(0.018)	(0.021)
Time FE	Yes	Yes
Baseline controls	Yes	Yes
Admit	Yes	No
Testing tendency	No	Yes
Observations	11,404	11,404

Notes: This table reports the 2SLS estimates of the effect of batching on patient outcomes, using specifications that include the additional controls of admitting status and testing tendency. $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

References

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

Gordon B. Dahl, Andreas Ravndal Kostøl, and Magne Mogstad. Family welfare cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752, November 2014. doi: 10.1093/qje/qju019. URL https://doi.org/10.1093/qje/qju019.

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

Sarah Eichmeyer and Jonathan Zhang. Pathways into opioid dependence: Evidence from practice variation in emergency departments. *American Economic Journal: Applied Economics*, 14(4):271–300, 2022. doi: 10.1257/app.20210048.