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 @doyle2015measuring, @dobbie2018effects, and @eichmeyer2022pathways. 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:

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

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.

First Stage

We document that the Mayo Clinic ED physicians exhibit wide, systematic variation in their propensity to batch order diagnostic tests. Tables 1 and 2 present the "first stage" results using both linear and probit regression models. In the linear model, being assigned to a physician with a 10 percentage point (pp) higher batch-tendency is associated with a 10.78 pp increase in the likelihood of having tests batch-ordered in the ED. The probit model suggests an even stronger relationship, with a 10 pp increase in batch-tendency associated with approximately a 35.9 pp increase in the probability of batching on average. Both models demonstrate a strong and statistically significant relationship between physician batch-tendency and the likelihood of test batching, supporting the relevance of our instrument.

Table 1: First Stage: Linear Regression of Batch Tendency on Test Batching

	Model 1	Model 2	Model 3
Batching Tendency	1.096***	1.078***	1.078***
	(0.034)	(0.027)	(0.027)
N	46,783	46,783	46,783
R^2	0.012	0.064	0.064
Adjusted R^2	0.011	0.061	0.061

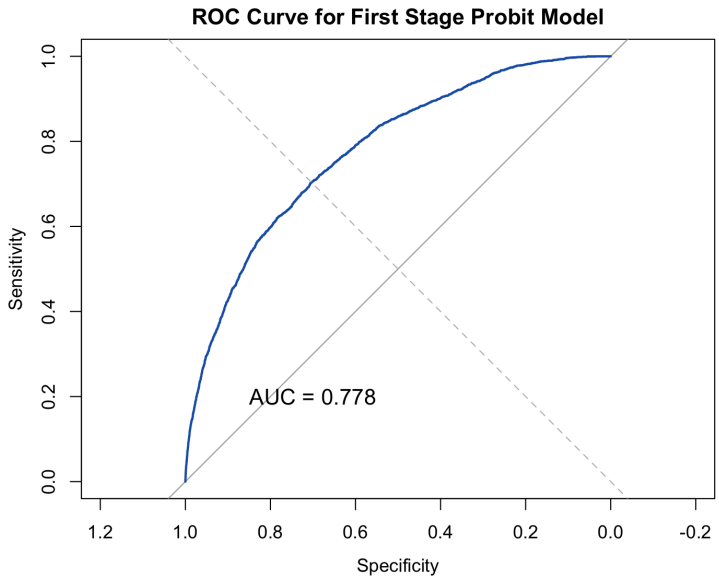
Linear regression estimates of the first stage. Robust standard errors are in parentheses. Model 1 includes only the batching tendency. Model 2 adds seasonality, shift fixed effects, and chief complaint. Model 3 further adds patient observables.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 2: First Stage: Probit Regression of Batch Tendency on Test Batching

	Model 1	Model 2	Model 3
Batching Tendency	8.507***	8.989***	8.984***
	(0.393)	(0.415)	(0.415)
N	46,783	46,783	46,783
Log Likelihood	-9,565.239	-8,617.761	-8,614.765
Akaike Inf. Crit.	19,210.480	17,509.520	17,515.530

Probit regression estimates of the first stage. Standard errors are in parentheses. Model specifications are the same as in Table 1.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Both the linear and probit models demonstrate the strength of our instrument, with F-statistics well above conventional thresholds in the linear model and highly significant coefficients in both models. The probit model, while potentially more theoretically appropriate for our binary outcome, yields results consistent with the linear model in terms of the strong, positive relationship between physician batch-tendency and the likelihood of test batching. These results support the relevance of our instrument in the subsequent instrumental variable analysis.



Reduced Form

Not controlling for testing inclination:

Table 3: Reduced Form

	Return and Admit (72hr)	Log ED LOS	Imaging Tests
	(1)	(2)	(3)
batch.tendency	-0.077^{***}	1.976^{**}	3.444^{***}
	(0.016)	(0.955)	(0.410)
N	46,783	46,783	46,783
R^2	0.005	0.302	0.257
Adjusted R^2	0.002	0.301	0.255
Residual Std. Error	0.115 (df = 46645)	0.472 (df = 46722)	0.666 (df = 46645)

Notes:

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

Controlling for testing inclincation:

Table 4: Reduced Form

	Return and Admit (72hr)	Log ED LOS	Imaging Tests
	(1)	(2)	(3)
batch.tendency	-0.091^{***}	-0.804	0.201^{**}
	(0.035)	(2.339)	(0.097)
test.inclination	0.004	0.850	0.972^{***}
	(0.011)	(0.596)	(0.019)
N	46,783	46,783	46,783
R^2	0.005	0.339	0.260
Adjusted R^2	0.002	0.337	0.258
Residual Std. Error (df = 46644)	0.115	0.459	0.665

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

Instrumental Variables Results

In this section, we examine the causal effects of batch ordering in the ED using a two-stage instrumental variables approach with a probit model in the first stage. This method, sometimes referred to as a probit-2SLS or probit-IV approach[1][2], allows us to account for the binary nature of our endogenous variable (batch ordering) while addressing potential endogeneity concerns.

Our first stage employs a probit model to estimate the probability of batch ordering as a function of our instrument (physician batch-tendency) and other covariates. We then use the predicted probabilities from this probit model as an instrument in the second stage linear regression.

Table 5 presents the effects on our main outcomes. We begin by estimating the effects using this probit-2SLS approach. Following this, we extend our analysis to estimate the direct and indirect effects of batch ordering on LOS using a causal mediation framework, which incorporates our probit first-stage estimates. To conduct the mediation analysis, we employ a custom bootstrapped approach that respects the clustered nature of our data. This method allows us to decompose the total effect of batch ordering on length of stay (LOS) into its direct effect and its indirect effect through the number of ED tests ordered. Our approach involves the following steps:

- We first use a probit model to predict the probability of batch ordering.
- We then estimate the effect of these predicted probabilities on the mediator (number of ED tests).
- Finally, we estimate the effects of both the predicted batching probabilities and the mediator on the outcome (LOS).
- We use clustered bootstrapping to estimate confidence intervals and p-values for the direct and indirect effects, ensuring that we account for the clustered structure of our data at the ED provider level.

This approach allows us to quantify not only the overall impact of batch ordering on LOS, but also to understand the mechanism through which this effect occurs, particularly the role played by the number of

[1] Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press. (Chapter 15 discusses nonlinear models in IV estimation)

[2] Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press. (They discuss various IV approaches, including with binary endogenous variables)

tests ordered. By using clustered bootstrapping, we ensure that our standard errors and confidence intervals accurately reflect the dependency structure within our data.

Table 5: IV Probit Regression Results

	Log ED LOS	Imaging Tests	Return and Admit (72hr)
Coefficient	-1.078257	0.6579455	-0.008485203
Std. Error	0.3353349	0.2127876	0.01395781
p-value	0.001303237	0.001989105	0.543245
R^2	0.3413561	0.5555298	0.005026105

Notes: Coefficients are from the second stage of an IV probit regression. Standard errors are clustered by ED_PROVIDER.
*** p < 0.01, ** p < 0.05, * p < 0.1.

Table 6: Mediation Analysis Results: Effect of Batch Ordering on Length of Stay

Effect	Estimate	95% CI	Z-score	P-value
Indirect (via ED tests)	0.134	[0.067, 0.180]	59.950	<0.001
Direct	-0.957	[-1.485, 0.347]	-2.805	0.005
Total	-0.823	–	–	–

Note: CI = Confidence Interval. Total effect is the sum of indirect and direct effects.

Our mediation analysis reveals a nuanced relationship between batch ordering and ED length of stay (LOS). Batch ordering has a significant positive indirect effect on LOS (0.134, p<0.001) mediated through the number of ED tests, suggesting that batching leads to more tests being ordered, which increases LOS. However, this is counterbalanced by a larger negative direct effect (-0.957, p=0.005), indicating that batch ordering also introduces efficiencies that reduce LOS. The total effect (-0.823) suggests that, overall, batch ordering decreases LOS. These findings imply that while batch ordering may increase the number of tests, it ultimately leads to shorter ED stays, likely due to improved process efficiency.

Table 7: Mediation Analysis Results: Effect of Batch Ordering on 72-Hour Return and Admission

Effect	Estimate	95% CI	Z-score	P-value
Indirect (via ED tests)	-0.00161	[-0.00219, -0.00036]	-9.028	<0.001
Direct	0.00011	[-0.01740, 0.01826]	0.006	0.995
Total	-0.00150	–	–	–

Note: CI = Confidence Interval. Total effect is the sum of indirect and direct effects.

The mediation analysis for 72-hour return and admission rates reveals a small but significant indirect effect of batch ordering. Batch ordering has a negative indirect effect (-0.00161, p<0.001) on 72-hour returns and admissions, mediated through the number of ED tests. This suggests that the increased testing associated with batch ordering slightly reduces the likelihood of patient returns or admissions within 72 hours. The direct effect (0.00011, p=0.995) is not statistically significant, indicating no evidence of a direct impact of batch ordering on this outcome. The total effect (-0.00150) suggests a minor overall reduction in 72-hour returns and admissions associated with batch ordering, primarily driven by the indirect effect through increased testing.