**Title**: Batch Ordering Imaging Tests in the Emergency Department and the Impact on Care Delivery

**Authors**: Jacob Jameson, MS1,2, Soroush Saghafian, PhD, MS2, Robert S. Huckman, PhD3, Nicole Hodgson, MD4

**Affiliations:** 1Interfaculty Initiative in Health Policy, Harvard University, Cambridge, MA, USA,2Harvard Kennedy School, Harvard University, Cambridge, MA, USA, 3Harvard Business School, Harvard University, Boston, MA, USA, 4Mayo Clinic, Department of Emergency Medicine, Phoenix, AZ, US

**Corresponding Author**: Jacob Jameson, MS,

Harvard Kennedy School, Mailbox 37, 79 JFK Street, Cambridge, MA 02138

Phone: 475-777-9252, Email: [jacobjameson@g.harvard.edu](mailto:jacobjameson@g.harvard.edu)

Manuscript Word Count 2901/3000

**Key Points**

Question: How does a physician's revealed preference towards batch ordering diagnostic imaging tests impact patient length of stay, returns, and cumulative testing in the Emergency Department?

Findings: In our study of 43,299 Emergency Department patient encounters at the Mayo Clinic of Arizona, a physician's inclination towards batch ordering diagnostic imaging tests was associated with increased length of stay, decreased probability of a 72-hour return, and increased patient testing.

Meaning: This study found that a test ordering strategy known as "batching," where a physician orders several tests at once rather than waiting on the results of one test before ordering another, has implications for patient outcomes and healthcare costs.

**Abstract**

Importance: Prior research has demonstrated that test ordering can significantly impact Emergency Department throughput, cost, and quality of care, yet there is no consensus on the optimal test ordering strategy.

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.

Design, Setting, and Participants: In this retrospective study, we use comprehensive EMR data from patients who visited the Mayo Clinic of Arizona Emergency Department between 10/6/2018 and 12/31/2019. During the study period, the ED recorded approximately 50,836 visits, managed across 26 treatment rooms and up to 9 hallway spaces.

Main Outcomes and Measure: Patient length of stay in the Emergency Department (measured in minutes and log-transformed), number of diagnostic imaging tests ordered during a patient encounter, and patient return to the Emergency Department within 72 hours. The association between outcomes and physician batch tendency was measured using a multivariable linear regression controlling for covariates necessary to achieve quasi-random assignment for patients to physicians.

Results: 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 0.065 (95% CI = [0.011,0.124], p < 0.001). Suggesting that having a physician with a batch tendency 1SD greater than the average physician was associated with an average 6.5% increase in total length of stay. We also find that a batch tendency 1SD greater than the average physician is associated with a 0.3 percentage point decrease in the probability of a 72-hour return rate, indicated by a coefficient of -0.003 (95% CI = [-0.005, -0.001], p < 0.001), implying that batching may lead to more comprehensive initial evaluations, reducing the need for short-term revisits. There is a notable association with an increased number of distinct imaging tests ordered, as evidenced by a coefficient of 0.096 (95% CI = [0.077, 0.114], p < 0.001), underscoring 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 ordering the next.

Conclusions and Relevance: This study highlights the considerable impact of physicians' diagnostic test ordering strategies on ED efficiency and patient care. These findings indicate that, on average, sequential ordering of tests enables physicians to serve patients more efficiently using the information obtained from prior tests (an information gain advantage). The results also highlight the need to develop guidelines to optimize ED test ordering practices.

**Introduction**

Emergency Departments (EDs) serve as critical junctures in healthcare delivery, balancing the immediate needs of patients with the systemic 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 safety1. 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 health2.

One of the least scrutinized aspects of this efficiency equation is the practice of batch ordering imaging tests. Given the long turnaround times of imaging tests, by placing multiple orders at once, the physician can ensure that their patient is in simultaneous waiting queues for each specific test3. While ostensibly a measure to expedite patient diagnosis and treatment, batch ordering intersects with several critical issues: the risk of over-testing, unnecessary radiation exposure, and the escalation of direct and indirect healthcare costs4,5,6. 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, and the impact on the patient's length of stay and overall healthcare costs.

Furthermore, the financial implications extend beyond the cost of the tests themselves. While 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 system7 This delicate balance between ensuring rapid, accurate diagnosis and minimizing unnecessary interventions is a central challenge in hospital management, reflecting broader concerns about the sustainability of healthcare practices8.

Despite its significance, the impact of batch ordering on these dimensions remains underexplored. The assumption that batch ordering represents an efficient 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 not only the length of stay and total testing volume but also considers the broader costs associated with this practice, both financial and in terms of patient safety.

By situating this investigation within the context of hospital management challenges, this research contributes to a nuanced understanding of ED operational efficiency. It seeks to unravel 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. Through this lens, the study addresses a pivotal yet understudied problem in hospital management, advocating for a more informed approach to diagnostic testing in the ED9,10.

**Methods**

**Study Design and Setting**: Our retrospective observational study was conducted in the Emergency Department (ED) of the Mayo Clinic of Arizona. During the study period, the ED recorded approximately 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 rotational patient-to-physician assignment system, which reduces systematic differences in patient populations served by different physicians.

We conducted a retrospective review of comprehensive ED operational data from 10/6/2018 through 12/31/2019, coinciding with initiating 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. We further restricted our sample to patient encounters serviced by full-time physicians and broad chief complaint areas seen in over 1,000 encounters over the study period (i.e., excluding rare complaints). The final sample was 43,299 patient encounters and contained no missing data concerning covariates used in the analysis.

**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 analyses11. In essence, barring arrival time and shift-level variation, the physician-to-patient matching can be deemed random. Table 1 in the Results section confirms that complaint and severity are balanced across physicians.

**Measurements:** 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 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 occurring when multiple diagnostic imaging tests are ordered within a 5-minute window. Sensitivity analyses around this cutoff point eTable 1 showed that our results are robust to this definition. Each imaging test (e.g., X-ray, CT scan) is considered a separate, distinct test for our study. Therefore, a batch in our study consists of two or more distinct imaging tests that differ in modality.

**Statistics Analysis**: To assess the impact of batching on various outcomes of interest, we developed a measure to quantify each physician's tendency to batch. This "batch tendency" score is a crucial element in our analysis, allowing us to explore the associations between batching behavior 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 estimate the following multivariable regression for each patient encounter:

Where is a dummy variable equal to one if patient had their imaging tests batch ordered on the encounter that took place on date . Fixed effects include year-month fixed effects, , 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, . Chief complaint by severity fixed effects, , were also included to increase precision. As stated earlier, these controls are 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, .

Then, for physician serving patient , we compute physician leave-one out residualized average (the leave-one out mean of for each physician) by excluding the current patient from the calculation and including all other patients served by physician during the study period. This leave-one-out measure effectively eliminates the mechanical bias resulting from patient own case influencing the physician's batch tendency score and captures the physician's general tendency to batch imaging tests across a wide range of cases12,13. Therefore, we can consider the physician's tendency to batch order image tests entirely independent of patient encounter characteristics.

After calculating each physician's leave-one-out batch rate, we z-score the batch rate into a uniform scale, facilitating more straightforward interpretation and comparison across physicians. A higher z-score indicates a greater propensity for batching compared to peers, while a lower score indicates a lesser tendency. Figure 1 shows the relationship between batch tendency and batch ordering at a specific patient encounter. This strong relationship between batch tendency and batch ordering allows us to think of batch tendency as an Instrumental Variable (IV), which addresses the problem of endogeneity in studying the impact of batching14. This will enable us to use batch tendency as a proxy for batching itself. The batch tendency also satisfies the exclusion restriction for a valid IV because we would not expect batch tendency to impact our primary outcomes in any other manner than through its effect on batching.

All statistical analyses were performed using R (version 4.3.2). All multivariable linear regression models control for month and time-of-day fixed effects, which is necessary to achieve quasi-random assignment. We additionally control for patient chief complaint and severity, an indicator for whether laboratory tests were ordered for the patient, and hospital factors, such as occupancy, to improve precision. We use robust standard errors clustered at the physician level to account for the assignment process of patients to physicians.

We were particularly interested in evaluating the influence of physicians' batch ordering tendency on three patient-level dependent variables: length of stay (LOS), the 72-hour return rate, and the number of distinct imaging tests ordered. Additional analyses involved examining interactions between batch tendency and other key variables, such as patient complaint and ESI, to explore whether the effect of batch ordering varied across different patient acuities and complaints. Because our data regarding 72-hour returns are limited to returns to the same ED, we should expect that the magnitude of our estimate is biased towards the null.

**Data Manipulation:** As evidenced in the literature, transforming the outcome variable can improve the performance of regression models. For right-skewed outcomes, such as the length of stay (LOS) in our dataset, applying a natural log transformation can lead to a more symmetric distribution and mitigate the influence of outliers15,16. This approach is somewhat analogous to count data models like Poisson or negative binomial regression, which are designed to handle skewed distributions inherent in count data and have been done by others17. As demonstrated in eFigure 1, the distribution of LOS is highly right-skewed; thus, we 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 2, which can be interpreted as an percent change in LOS for a given 1 unit increase in our independent variable of interest, where is the coefficient on our independent variable of interest.

**Results**

Table 1 displays the results of a Wald balance test, showing that complaints and severity of patient encounters are balanced across physicians in our study's cohort. In other words, due to the random assignment, all physicians served a similar portfolio of patients regarding presenting complaints and severity. This is a critical aspect of our study, ensuring that differences in test ordering behavior are attributable to physician practice patterns rather than patient characteristics.

The data also indicate differences in test ordering practices across complaint categories (Figure 2). 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 approximately 31,498 out of the 43,299 patient encounters during our study period. While only 2,421 of these encounters involved image batching (7.7%), we note that 7,181 of the non-batched encounters resulted in the physician ordering at least one more imaging test after placing the first order (22.8%).

Table 2 presents the linear regression coefficients for the impact of batch tendency on three primary outcomes: the natural logarithm of ED length of stay (ln(LOS)), the 72-hour return rate, 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 0.065 (95% CI = [0.011,0.124], p < 0.001). Suggesting that having a physician with a batch tendency 1SD greater than the average physician results in an average 6.5% increase in total length of stay. Conversely, batch tendency 1SD greater than the average physician is associated with a 0.3 percentage point decrease in the probability of a 72-hour return rate, indicated by a coefficient of -0.003 (95% CI = [-0.005, -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 0.096 (95% CI = [0.077, 0.114], p < 0.001), underscoring 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 ordering the next.

Figure 3 displays the results of the subgroup analysis stratified by the patient's Emergency Severity Index (ESI) and broad patient complaint category (as defined in eTable 2 and eTable 3). Results indicate heterogeneity in the effect of batch tendency across patient complaints and acuity. Notably, among the most acute patients (ESI 1 and 2), the propensity to batch order image tests was generally associated with significant increases in LOS and the total number of imaging tests ordered. We do not see substantial reductions in the probability of a 72-hour return (though this coefficient is biased towards the null), except for the case of patients presenting with cardiac/chest-related complaints. This may imply that a more comprehensive testing approach, namely one associated with a greater testing volume, decreases the likelihood of short-term returns for patients within this subgroup18.

**Discussion**

The patterns of diagnostic test ordering in the Emergency Department (ED) have profound implications on the efficiency of care delivery and patient outcomes, as our study critically highlights. This investigation into the variances in physician test ordering behaviors within a controlled ED environment brings to light the need for targeted diagnostic strategies over a one-size-fits-all approach. Our findings resonate with the broader discourse on ED management, where evidence-based medicine is increasingly sought to enhance patient care and system efficiency19.

The inclination towards batching or non-batching test orders among physicians—despite uniform protocols—raises fundamental questions about the underpinnings of clinical decision-making. Notably, our study revealed that non-batchers, who potentially employ a more judicious and sequential approach to ordering tests, could achieve shorter patient length of stay (LOS) without negatively impacting the 72-hour return rates, a surrogate for quality of care. This is due to the information gain advantage of sequential test ordering. These insights align with previous research emphasizing the importance of tailored diagnostic pathways in achieving optimal health outcomes and operational efficacy20,21.

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 findings22. Moreover, the economic implications are substantial, with the overuse of diagnostic tests contributing significantly to the escalating costs of healthcare23. As such, our results are a clarion call for interventions to modulate physicians' batching behaviors to align more closely with precision medicine principles.

Implementing a more individualized approach to patient-physician assignment in the ED is complex. Recent initiatives have experimented with predictive algorithms to optimize patient-physician matching based on various factors, including patient complaints and physician expertise24. Incorporating physicians' test ordering tendencies could further refine these algorithms, potentially enhancing patient satisfaction and outcomes while curtailing unnecessary resource utilization21.

Future studies should investigate the subtleties of information gained 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. Investigations following our study should aim to quantify the causal impact of targeted testing strategies on a broader scale across diverse healthcare systems and patient populations, ensuring the generalizability of the recommended practices.

**Limitations**

Our study involves multiple considerations that may limit the interpretation and application of our findings. While we accounted for the random assignment of patients to physicians, the variation we observe across physicians could stem from myriad sources, including physicians' training, accumulated experience, and general inclinations toward more testing25. 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, while we consider ED physicians to be independent actors, it is known that they affect each other's speed and quality26. Therefore, moving beyond associative insights is imperative as research in this area of inquiry progresses. We also acknowledge the possibility of type I error due to multiple comparisons in our subgroup analysis, and these analyses should be taken as exploratory.

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.

**Conclusion**

Our study contributes to a critical conversation on optimizing diagnostic processes in the ED and underscores the need for individualized diagnostic strategies to enhance operational efficiency and patient care. Through a detailed investigation of batch versus non-batch ordering practices, we have highlighted the profound implications of these behaviors on patient length of stay, resource utilization, and hospital costs, providing new evidence that challenges the current paradigm of diagnostic testing in the ED.

Notably, batching was associated with increased ED lengths of stay and testing, with a small but significant decrease in the probability of a 72-hour return, suggesting that a more discerning approach to test ordering can be both efficient and possibly beneficial to patient outcomes. These findings echo the call for precision medicine principles to be integrated into ED operations, emphasizing the importance of tailoring diagnostic processes to individual patient needs to avoid over-testing pitfalls, such as unnecessary radiation exposure and the financial burden of healthcare delivery.

**Conflicts of Interest**: There are no conflicts of interest to report.

**References**

1. Cournane S, Conway R, Creagh D, Byrne DG, Sheehy N, Silke B. Radiology imaging delays as independent predictors of length of hospital stay for emergency medical admissions. Clin Radiol. 2016;71(9):912-918. doi:10.1016/j.crad.2016.03.023
2. Darraj A, Hudays A, Hazazi A, Hobani A, Alghamdi A. The Association between Emergency Department Overcrowding and Delay in Treatment: A Systematic Review. Healthcare (Basel). 2023;11(3):385. Published 2023 Jan 29. doi:10.3390/healthcare11030385
3. Perotte R, Lewin GO, Tambe U, et al. Improving Emergency Department Flow: Reducing Turnaround Time for Emergent CT Scans. AMIA Annu Symp Proc. 2018;2018:897-906. Published 2018 Dec 5.
4. Lyu H, Xu T, Brotman D, Mayer-Blackwell B, Cooper M, et al. Overtreatment in the United States. PLOS ONE. 2017;12(9):e0181970. doi:10.1371/journal.pone.0181970
5. Jain S. Radiation in medical practice & health effects of radiation: Rationale, risks, and rewards. J Family Med Prim Care. 2021;10(4):1520-1524. doi:10.4103/jfmpc.jfmpc\_2292\_20
6. Tamburrano A, Vallone D, Carrozza C, et al. 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. 2020;15(8):e0237159. Published 2020 Aug 6. doi:10.1371/journal.pone.0237159
7. Sartini M, Carbone A, Demartini A, et al. Overcrowding in Emergency Department: Causes, Consequences, and Solutions-A Narrative Review. Healthcare (Basel). 2022;10(9):1625. Published 2022 Aug 25. doi:10.3390/healthcare10091625
8. Ibanez MR, Clark JR, Huckman RS, Staats BR. Discretionary Task Ordering: Queue Management in Radiological Services. Manag Sci. 2017;64(9):4389-4407.
9. Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL. Patient Streaming as a Mechanism for Improving Responsiveness in Emergency Departments. Oper Res. 2012;60(5):1080-1097.
10. Saghafian S, Austin G, Traub SJ. Operations Research/Management Contributions to Emergency Department Patient Flow Optimization: Review and Research Prospects. IIE Trans Healthc Syst Eng. 2015;5(2):101-123.
11. Traub SJ, Stewart C, Didehban R, et al. Emergency Department Rotational Patient Assignment. Ann Emerg Med. 2016;67(2):206-215.
12. Eichmeyer S, Zhang J. Pathways into Opioid Dependence: Evidence from Practice Variation in Emergency Departments. Am Econ J Appl Econ. 2022;14(4):271-300.
13. Dobbie W, Goldin J, Yang CS. The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. Am Econ Rev. 2018;108(2):201-240.
14. Konetzka RT, Yang F, Werner RM. Use of instrumental variables for endogenous treatment at the provider level. Health Econ. 2019;28(5):710-716. doi:10.1002/hec.3861.
15. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analyzing health care utilization and costs. Annu Rev Public Health. 1999;20:125–144. doi:10.1146/annurev.publhealth.20.1.125.
16. Cots F, Elvira D, Castells X, Sáez M. Relevance of outlier cases in case mix systems and evaluation of trimming methods. Health Care Manag Sci. 2003;6(1):27–35. doi:10.1023/A:1021908220013.
17. Verburg IWM, de Keizer NF, de Jonge E, Peek N. Comparison of regression methods for modeling intensive care length of stay. PLoS ONE. 2014;9(10):e109684. doi:10.1371/journal.pone.0109684.
18. Carpenter CR, Raja AS, Brown MD. Overtesting and the Downstream Consequences of Overtreatment: Implications of "Preventing Overdiagnosis" for Emergency Medicine. Acad Emerg Med. 2015;22(12):1484-1492. doi:10.1111/acem.12820
19. Masic I, Miokovic M, Muhamedagic B. Evidence based medicine - new approaches and challenges. Acta Inform Med. 2008;16(4):219-225. doi:10.5455/aim.2008.16.219-225
20. Singh H, Graber ML. Improving Diagnosis in Health Care--The Next Imperative for Patient Safety. N Engl J Med. 2015;373(26):2493-2495. doi:10.1056/NEJMp1512241
21. Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL. Complexity-augmented triage: A tool for improving patient safety and operational efficiency. Manuf Serv Oper Manag. 2014;16(3):329-345.
22. Lam JH, Pickles K, Stanaway FF, Bell KJL. Why clinicians overtest: development of a thematic framework. BMC Health Serv Res. 2020;20(1):1011. Published 2020 Nov 4. doi:10.1186/s12913-020-05844-9
23. Müskens JLJM, Kool RB, van Dulmen SA, Westert GP. Overuse of diagnostic testing in healthcare: a systematic review. BMJ Qual Saf. 2022;31(1):54-63. doi:10.1136/bmjqs-2020-012576
24. Atkinson MK, Saghafian S. Who should see the patient? On deviations from preferred patient-provider assignments in hospitals. Health Care Manag Sci. 2023;26:165–199.
25. Abaluck J, Agha L, Kabrhel C, Raja A, Venkatesh A. The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care. Am Econ Rev. 2016;106(12):3730-64.
26. Saghafian S, Imanirad R, Traub SJ. Do Physicians Influence Each Other's Performance?: Evidence from the Emergency Department. Harvard Kennedy School, John F. Kennedy School of Government Working Paper Series. 2019.

**Figure Legend**

Table 1: Balance Test for Random Assignment of Patients Included in Analytic Sample (N=43,299)

Table 2: Multivariable Regression Results of Primary Outcomes on Batch Tendency

Figure 1: Relevance of Batch Tendency on Batch Ordering Probability

Figure 2: Variation in Physician Imaging Batch Rates

Figure 3: Regression Coefficients with Confidence Intervals from Subgroup Analysis

**Table 1: Balance Test for Random Assignment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Chief Complaints** | **Frequency No. (%)** | **F-Statistic** | ***p-value*** |
| Abdominal Complaints | 6232 (14%) | 2.587 | 0.108 |
| Back or Flank Pain | 2552 (6%) | 1.637 | 0.201 |
| Chest Pain | 3525 (8%) | 0.407 | 0.524 |
| Extremity Complaints | 5265 (12%) | 1.847 | 0.174 |
| Assaults and Trauma | 2381 (5%) | 0.023 | 0.880 |
| Gastrointestinal Issues | 3323 (8%) | 0.105 | 0.746 |
| Neurological Issue | 3495 (8%) | 0.135 | 0.713 |
| Shortness of Breath | 2966 (7%) | 1.324 | 0.250 |
| Skin Complaints | 2178 (5%) | 0.383 | 0.536 |
| Upper Respiratory Symptoms | 1917 (4%) | 0.017 | 0.896 |
| **Emergency Severity Index (ESI)** | **Frequency No. (%)** | **F-Statistic** | ***p-value*** |
| ESI 1 or 2 | 13914 (32%) | 0.011 | 0.915 |
| ESI 3, 4, or 5 | 29386 (68%) | 0.010 | 0.921 |
| **Vital Signs** | **Frequency No. (%)** | **F-Statistic** | ***p-value*** |
| Tachycardic | 8367 (19%) | 0.118 | 0.731 |
| Tachypneic | 4003 (9%) | 0.043 | 0.836 |
| Febrile | 1021(2%) | 0.936 | 0.333 |
| Hypotensive | 647 (1%) | 1.127 | 0.288 |

*Table 1 reports the results of a Wald test, which was conducted to assess the balance of chief complaints across physicians in our dataset. We created chief complaint categories before analysis by grouping similar presenting issues. Vital signs were categorized as follows: tachycardia (pulse more significant than 100), tachypnea (respiratory rate greater than 20), fever (temperature greater than 38∘ C), and hypotension (systolic blood pressure less than 90). A balanced distribution implies that complaints and severity are evenly distributed across physicians, which we expect to be the case due to randomization. The Wald F-statistic and p-value are reported. Robust standard errors (type HC1) accounted for potential heteroscedasticity in the data.*

**Table 2: Main Multivariable Regression Results of Primary Outcomes on Batch Tendency**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Dependent Variables | | | |
|  | |  |  |  |
|  | | ln(LOS) | 72 Hour Return | Number of Distinct Imaging Tests |
| Batch Tendency | | 0.065\*\*\*  (0.023) | -0.003\*\*\*  (0.001) | 0.096\*\*\*  (0.009) |
|  | |  |  |  |
| Controlling for time and shift? | | Yes | Yes | Yes |
| Controlling for complaint and ESI? | | Yes | Yes | Yes |
| Controlling for hospital occupancy? | | Yes | Yes | Yes |
| Controlling for lab tests ordered? | | Yes | Yes | Yes |
|  | |  |  |  |
| Observations | | 43,299 | 43,299 | 43,299 |

*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), as well as patient level variables, hospital occupancy, and whether the patient also had laboratory tests ordered in their visit. Standard errors are clustered at the physician level.*

*\*p<0.1; \*\*p<0.05; \*\*\*p<0.01*

**Figure 1: Relevance of Batch Tendency on Batch Ordering Probability**

A graph with a red line

Description automatically generated

Figure 1 shows the predicted probability 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 x-axis represents the batch tendency score, which measures the physician's tendency to batch-order. The red line represents the predicted probability of batch-ordering at a specific patient encounter, and the shaded area represents the 95% confidence interval.

**Figure 2: Variation in Physician Imaging Batch Rates**

A graph with different colored dots

Description automatically generated

*Figure 2 illuminates the marked differences among physicians in their propensity to batch order imaging tests. The 24 physicians are represented with points, revealing that specific complaint areas have more variance than others when it comes to differing batch rates among physicians.*

**Figure 3: Regression Coefficients with Confidence Intervals from Subgroup Analysis**

A graph of a graph

Description automatically generated with medium confidence

*Each complaint, severity subgroup regression model was run using controlling for the same covariates in the main analysis.*