Strategic Test Ordering in the Emergency Department and the Impact on Care Delivery

Jacob Jameson, MS[1]

Soroush Saghafian, PhD, MS[2]

Robert S. Huckman, PhD[3]

Nicole Hodgson, MD[4]

Abstract

Study Objective: The study aimed to investigate the impact of emergency physicians' (EPs) diagnostic test ordering strategies—specifically, the tendency to batch versus order tests sequentially—on resource utilization, patient length of stay (LOS), and outcomes in an Emergency Department (ED) setting.

Methods: This retrospective analysis reviewed over 43,000 ED visits at the Mayo Clinic of Arizona. The focus was on the variability in EPs' test-ordering strategies, defining batching as ordering multiple diagnostic tests within a 5-minute window. The study assessed the relationship between batching tendencies and patient outcomes (LOS and resource utilization), adjusting for patient conditions and severity. Patients were randomly assigned to physicians to mitigate selection bias.

Results: There was wide variability in batching rates among physicians, with a 30-percentage point difference between the lowest and highest batching rates. Higher batching tendencies were associated with a 6.5% increase in LOS (95% CI: 2.0-11.2%) and 12 additional tests per 100 patient encounters (95% CI: 7.5-16.5), indicating significant higher resource utilization without observed improvements in patient outcomes, such as 72-hour return rates.

Conclusions: The study highlights the considerable impact of physicians' diagnostic test-ordering strategies on ED efficiency and patient care. Specifically, a higher tendency to batch diagnostic orders correlates with increased LOS and resource utilization without improving patient outcomes. 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 for developing guidelines to optimize ED test-ordering practices.

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

1. Introduction

Emergency Departments (EDs) sit at the forefront of healthcare delivery, often grappling with inefficiencies that impact patient outcomes and the broader health system. One potential determinant of these inefficiencies is variability in physicians' approaches to diagnostic test ordering. Drawing on data from over 43,000 ED visits at the Mayo Clinic of Arizona, this study reveals pronounced variability in physicians' tendencies to batch-order or sequentially-order diagnostic tests. Individual physicians have batching rates between 25% and 55% of their total patient encounters. This variation persists even among physicians practicing under the same guidelines within the same hospital, serving patients randomly assigned to them.

The data indicate that differences in test-ordering practices were not mere anomalies but reflect individual physician decision-making patterns consistent across complaint categories (Figure 1). Notably, physicians with a higher batching rate in one category of complaints tend to have a higher batching rate in other categories. Our findings underscore a potential imperative: designing more effective and efficient care delivery methods by refining guidelines that target test-ordering strategies. This is crucial in optimizing ED operations[1,2], curbing unnecessary expenditures, and enhancing patient outcomes. This study adds to the literature by shedding light on an overlooked aspect of ED practice, quantifying its impact on resource utilization and patient length of stay.

Figure 1

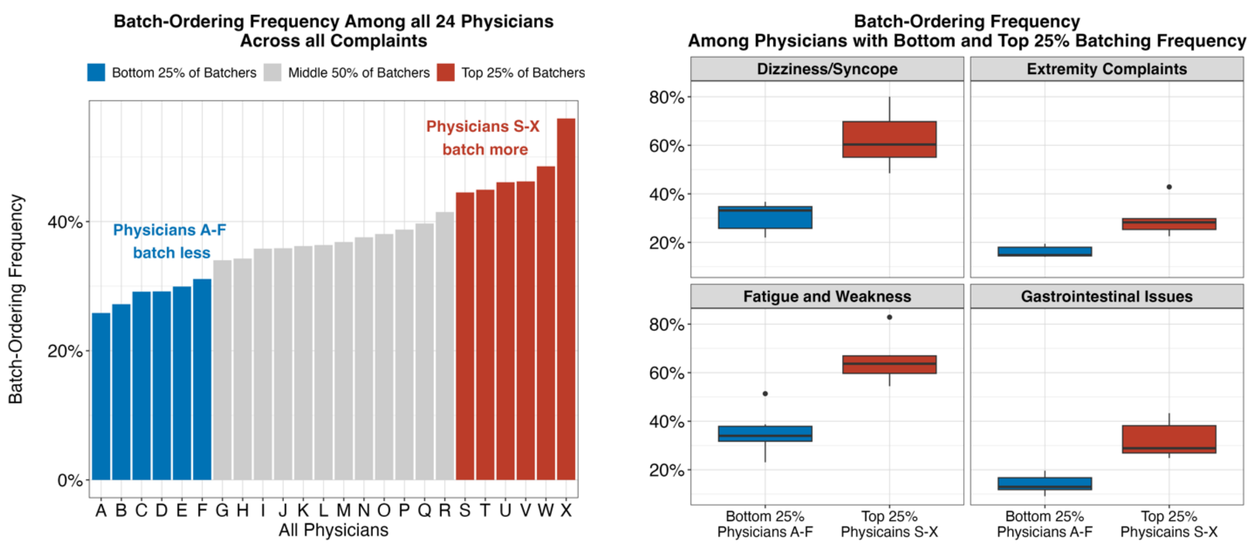

Figure 1 illuminates the marked differences among physicians in their propensity to batch-order diagnostic tests. Physicians are mapped on the x-axis, revealing those with a systematically heightened tendency to batch (in red) compared to their peers who batch less frequently (in blue). The quartile of batchers is calculated based on the batching rate across all complaint areas per physician. Physicians with a higher batching rate in one category of complaints tend to have a higher batching rate in other categories.

2. Study Data and Methods

Our 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 43,328 patient encounters.

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[3], which practically removes potential selection bias concerns from our analyses. In essence, barring arrival time and shift-level variation, the physician-to-patient matching can be deemed random. Table 1 displays the balance test results, which show that the main complaint and severity of patient encounters are indeed equitably distributed across physicians in our study's cohort. In other words, due to the random assignment, all physicians served a similar portfolio of patients. 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.

2.1 Definitions

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 tests are ordered within a 5-minute window. Sensitivity analyses around this cutoff point showed that our results are robust to this definition. This definition aligns with the concept of batching as a single comprehensive effort. For our analysis, all lab-based tests are categorized as one distinct type of testing (lab). In contrast, each imaging test (e.g., X-ray, CT scan) is considered a separate, distinct test. Therefore, a batch in our study consists of two or more distinct diagnostic tests, which could be a combination of lab and imaging tests or multiple imaging tests that differ in modality. This approach reflects the realistic diagnostic strategies in an ED setting and distinguishes between batching and non-batching behaviors. The differentiation between lab tests as one collective unit and imaging tests of different modalities as individual units is based on their operational impact in the ED, particularly regarding patient waiting time and processing queues.

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, for physician j serving patient i, we compute physician $j's$ batch rate (number of batched patient encounters out of their total patient encounters) by excluding the current patient i from the calculation and including all other patients served by physician j during the study period. This leave-one-out measure effectively eliminates the mechanical bias resulting from the patient $i's$ own case influencing the physician's batch tendency score[4]. Using this measure, we capture the physician's general tendency to batch across a wide range of cases rather than being skewed by any single patient encounter.

Table 1

Chief Complaints	**Frequency (%)**	**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
Falls, 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 (%)**	**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 (%)**	**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.

After calculating each physician's leave-one-out batch rate, we standardized these values by creating z-scores. This standardization process converts the batch tendency scores 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 2 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 batching[5]. This will enable us to use batch tendency as a proxy for batching itself. The batch tendency also provides a valid IV for studying the effect of batching because we would not expect batch tendency to impact outcomes of interest for a patient—such as number of tests ordered, LOS, or 72-hour rate of return—in any other manner than through its effect on batching (thereby satisfying the exclusion restriction for a valid IV).

3. Results

Results in Figure 3a indicate that, after controlling for patient conditions and severity, being seen by a physician with a batch tendency 1 SD greater than that of the average physician is associated with a 6.5% increase in LOS and an average of 12 excess tests per 100 patient encounters. These results indicate that being seen by a batcher may lead to efficiency losses on average. Furthermore, although batching results in extra tests and prolonged LOS, it does not yield improvements in patient outcome measures widely used by EDs as proxies for quality of care, such as the 72-hour rate of return (Figure 2a).

Figure 2

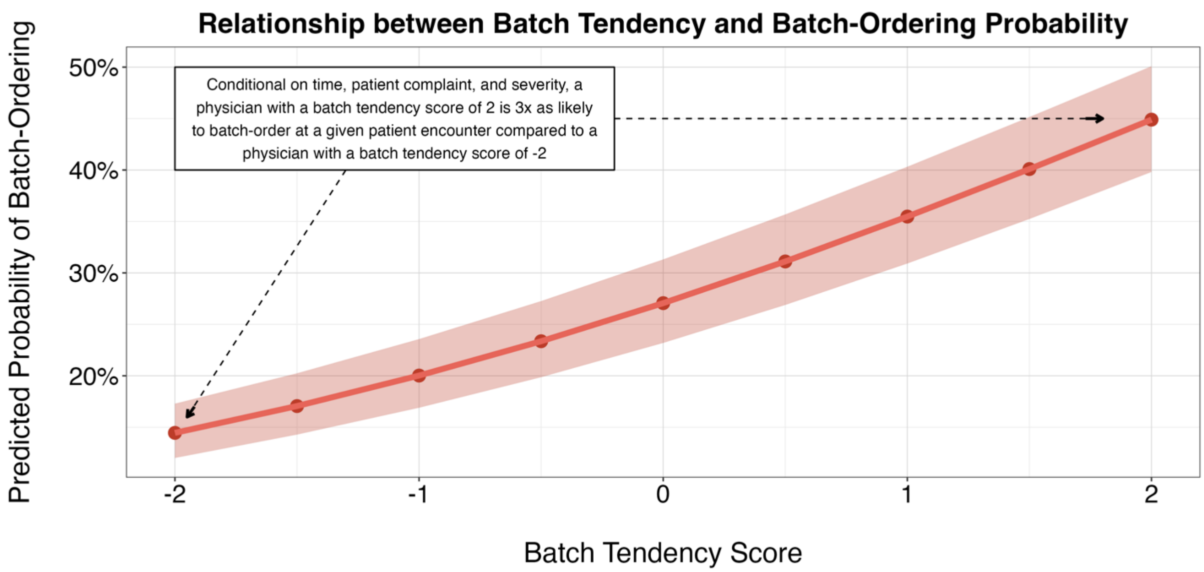

Figure 2 shows the predicted probability of batch-ordering at a given patient encounter, conditional on time, patient complaint, and severity, from a logistic regression model controlling for these features. The x-axis represents the batch tendency score, which measures the physician's tendency to batch-order. 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 3b displays the optimal test ordering strategy based on the patient's Emergency Severity Index (ESI), stratified by five broad patient complaint categories. Results indicate that, except for the most critical cases (ESI 1), sequentially ordering tests (i.e., non-batching) is generally preferred when the goal is to reduce the length of stay and the total number of diagnostic tests ordered. Nevertheless, there is apparent heterogeneity in optimal testing strategy by acuity, complaint, and the overall objective (i.e., decreased length of stay vs. decreased likelihood of 72-hour rate of return). The complaints and severity where batching is the optimal strategy are likely driven by scenarios where more diagnostic tests are generally required to care for the patient appropriately. Given the importance of optimizing ED operations and patient outcomes considering trade-offs in information gain, speed, and quality[6], we hope our findings motivate future research into determining when a physician should or should not batch order tests.

Figure 3

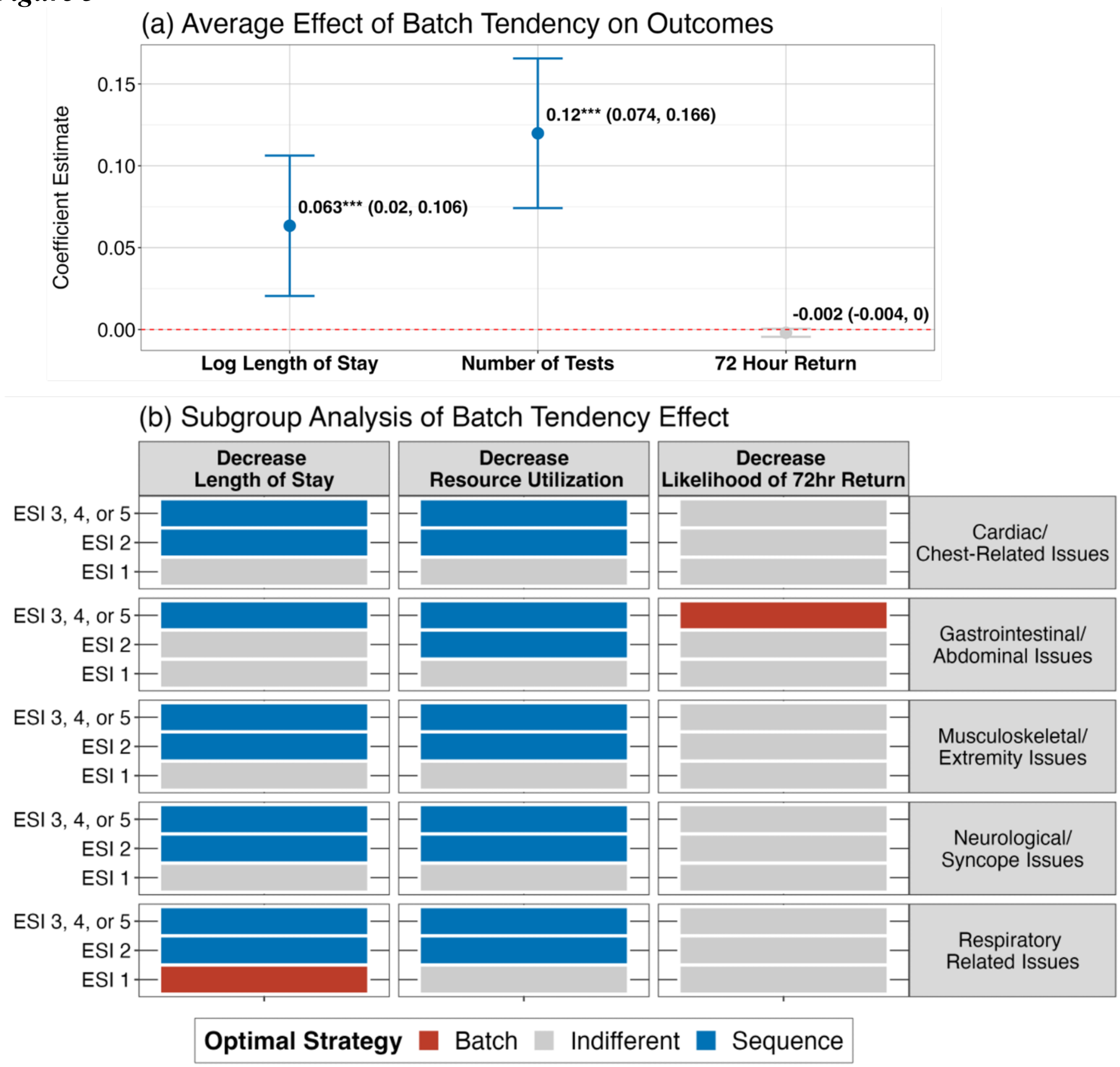

*Figure 3a displays the regression results controlling for Year-Month and Hospital-Day of week-Hour of day fixed effects. The chief complaint comes from the clean complaint the patient had at the initial encounter. Figure 3b displays the average marginal effects of batch tendency after controlling for the same fixed effects, but stratifying by complaint, and interacting with ESI. Robust standard errors are clustered at the physician level. *p<0.1; **p<0.05; ***p<0.01. Colors (blue or red) indicate that the average marginal effect of batch tendency was statistically significant (p<0.05) for a given ESI by Complaint.*

4. Limitations

The variation we observe across physicians could stem from myriad sources, including physicians' training, accumulated experience, ingrained diagnostic philosophies, and general inclinations toward more testing[7]. 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. For this and similar reasons, moving beyond associative insights is imperative as research in this area of inquiry progresses. While we have taken advantage of the random assignment in our setting, future research should continue exploring causality, dissecting the intricate relationship between batch ordering, potential over-testing, and their collective impact on overall efficiency and patient outcomes. Finally, while we considered ED physicians as independent actors, it is known that they affect each other's speed and quality[8]. Future studies can extend our results by investigating whether the batching tendency among physicians and its impact on speed and quality is driven by the influence of physicians on each other's practice.

5. Conclusion

The diagnostic process is central to ED operations and can significantly impact patient outcomes. Our investigation into diagnostic test ordering patterns reveals stark differences, even among physicians operating within identical environments and adhering to similar guidelines.

Our results raise the critical question of whether—if presented with a choice upon entering the ED—a given patient would be better served by a batcher or a non-batcher[9]. Regarding the preferred patient-physician assignment, the question of who should see the patient has received some attention in recent years in different settings[10]. However, this approach to preferred assignment has not been explored in the ED, where diagnostic testing is critical.

Central to our findings is the indication that not batching ED tests yields an information gain advantage, enabling physicians to reduce the number of tests and shorten the patient length of stay without any negative impact on patient outcomes such as 72-hour rate of return. Future studies, however, are required to validate our findings and obtain more insights into optimal testing strategies. Understanding the trade-off between this information gained from sequencing and the benefits that batching might offer is critical. Such understanding may help streamline ED operations and extend current approaches to improving ED efficiency and responsiveness[1,2,11], curtailing healthcare expenditures, and improving the quality of patient care.

References

1. 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.

2. 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.

3. Traub SJ, Stewart C, Didehban R, et al. Emergency Department Rotational Patient Assignment. Ann Emerg Med. 2016;67(2):206-215.

4. 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.

5. 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.

6. Saghafian S, Hopp WJ, Iravani SMR, Cheng Y, Diermeier D. Workload Management in Telemedical Physician Triage and Other Knowledge-Based Service Systems. Manag Sci. 2018;64(11):5180-5197.

7. 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.

8. 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.

9. Ibanez MR, Clark JR, Huckman RS, Staats BR. Discretionary Task Ordering: Queue Management in Radiological Services. Manag Sci. 2017;64(9):4389-4407.

10. 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.

11. 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.