Thinking versus Doing:
Cognitive Capacity, Decision Making and Medical Diagnosis

PRELIMINARY - DO NOT CITE OR CIRCULATE

Ben Handel
Louis-Jonas Heizlsperger
Jonas Knecht
Jonathan Kolstad
Ulrike Malmendier
Filip Matejka*

March 19, 2025

Abstract

The process of diagnosis in medicine depends on high dimensional, complex belief formation, repeated experimentation and data aggregation from multiple sources (e.g. labs, images, discussions with patients). In this paper we explore the role of point-in-time cognitive capacity on decision making of physicians. Using detailed audit log data from emergency rooms, we access how cognitive load during the shift affects future choices of tests as well as the treatment. We draw upon theories of inattention and information acquisition to infer how much is physicians' knowledge refined when making a particular choice. For instance, we find that high cognitive load implies that physicians make less informed choices in the future, perhaps choose to think less, and instead substitute this with more orders of diagnostic tests.

*Handel: UC Berkeley and NBER. Email: handel@berkeley.edu. Heizlsperger: UC Berkeley. Email: louis.heizlsperger@berkeley.edu. Kolstad: UC Berkeley and NBER. Email: jkolstad@berkeley.edu. Knecht: UC Berkeley. Email: jonasknecht@berkeley.edu. Malmendier: UC Berkeley and NBER. Email: ulrike@berkeley.edu. Matejka: CERGE-EI. Email: fmatejka@gmail.com. Helpful and much appreciated suggestions were provided by ... and seminar and conference participants at ...

1 Introduction

The process of medical care is intrinsically complex. Central to many, if not all, aspects of health care production is diagnosis. Physicians and other health care providers engage in an intricate process of information acquisition, synthesizing inputs from patient history, conversations, physical examinations, imaging, and laboratory tests. This information is then used to formulate and test diagnostic hypotheses under conditions of significant time constraints, high stakes, and institutional complexity.

At the core of this decision-making lies human cognition, where providers must integrate high-dimensional information to arrive at accurate diagnoses and subsequent treatment plans. The cognitive demands involved suggest that the availability of mental resources at any given moment may influence diagnostic accuracy and clinical choices. However, much of the existing literature on physician decision-making has been grounded in standard economic models or traditional behavioral frameworks, primarily focusing on external incentives, market structures, or cognitive biases as drivers of variation in care.

This paper takes a different approach by emphasizing cognitive capacity as a fundamental factor in medical decision-making. Arriving at a diagnosis requires mental effort, and fluctuations in cognitive resources may lead to variation in clinical judgments. This perspective contrasts with the rational-agent framework, which attributes differences primarily to external incentives, as well as with conventional behavioral models that emphasize systematic biases (e.g. probability misjudgments in prospect theory). Notably, neither approach readily explains why an identical patient presenting at different times might receive substantially different care.

Recent theoretical and experimental advances suggest that models of inattention, cognition and complexity can unify disparate behavioral decision-making frameworks. In the context of healthcare, such a unification is not merely theoretical but has direct implications for policy and organizational design. Traditional interventions—such as modifying reimbursement structures, enforcing clinical guidelines, or increasing training—may be less effective than commonly assumed if decision quality is constrained by cognitive capacity. Instead, interventions aimed at optimizing mental resources—whether through policy, organizational restructuring, job design

or the integration of decision-support technologies (e.g., artificial intelligence)—may hold significant potential for improving healthcare efficiency and outcomes. In contrast, a fragmented view of decision-making, grounded in setting-specific heuristics or physician-specific objective functions, presents greater challenges for both evaluation and intervention design. Empirically, studying the role of cognitive load in clinical decision-making presents significant challenges, particularly given that cognitive mechanisms are not directly observable. This limitation has led much of the literature in cognitive economics to focus on controlled laboratory settings, where cognitive load or task complexity can be systematically manipulated. However, medical diagnosis presents a unique case: the cognitive effort involved in forming a diagnosis ("thinking") can often be substituted with reliance on diagnostic tests ("doing"). Because the use of diagnostic tests is observable, it provides a valuable empirical window into the underlying cognitive mechanisms that precede testing.

To explore these dynamics, we develop a conceptual framework that formalizes the role of cognitive resources in diagnostic decision-making. Our framework not only contributes to the broader literature on physician behavior but also offers policy-relevant insights into the design of healthcare systems that account for the cognitive constraints of providers.

This paper argues that clinical decision-making must be understood within a model that accounts for the cognitive realities of medical practice. Specifically, ask whether the mental effort required to diagnose and treat one patient may be shaped by the complexity of other cases a physician encounters? Cognitive load is not compartmentalized; it interacts across decisions. If physicians allocate limited cognitive resources across cases, then the demands of some patients may spill over, influencing how others are treated.

2 Conceptual Framework

Physician faces a patient of an unknown state $d \in X$, where $X = \{1, ..., D\}$ is the set of all possible states. The state represents an unknown diagnosis of the patient, and $D > 1$ is the number of all possible diagnoses.

Let f be physician's subjective belief about a patient's state, i.e., f is a probability distribution

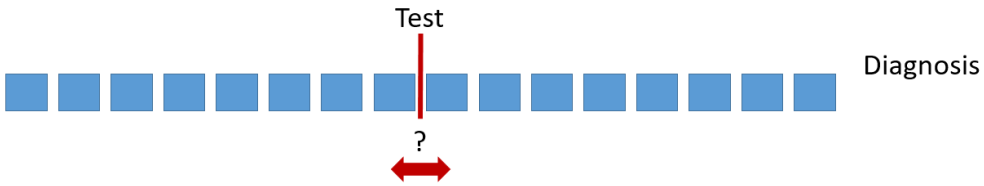


Figure 1: Uniform prior belief f^0, $D = 16$.

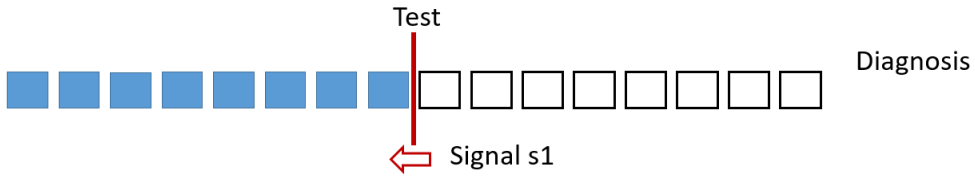


Figure 2: Belief f^1 after one signal with a realization is "bottom half".

on X. We denote f^0 to be a prior belief and f^t the be a belief after t signals are received. We assume that f^0 is uniform, i.e., $f^0(d) = 1/D$.

The physician chooses how to learn about state d and what treatment $y \in [1, ..., D]]$ to prescribe to maximize the following objective:

$$E\Big[-|y - d|\Big] - \text{cost of information.} \tag{1}$$

The first term represents a (health or resource) loss from misalignment between the chosen treatment and the patient's true state. Typically, the better information the physician possesses, the smaller this loss is - the physician is able to match the optimal treatment $y^* = d$ more precisely. The cost of information in the second term is described below. In depends on how much of her cognitive ability the physician deploys to the task and also how many diagnostic tests she uses to determine d.

Information acquisition The physician first acquires mental signals, i.e., thinks about the patient he faces, and then acquires signals from diagnostic tests that she orders.

All signals have a form of a binary partition. Given a current belief $f^t = U[a, b]$, the physician finds out whether the state is in a lower or an upper half of the states, i.e., f^{t+1} is either $U[a, \frac{a+b-1}{2}]$ or $U[\frac{a+b+1}{2}, b]$. See Figures 1 and 2. For simplicity, we here assume that D equals a power of 2 and

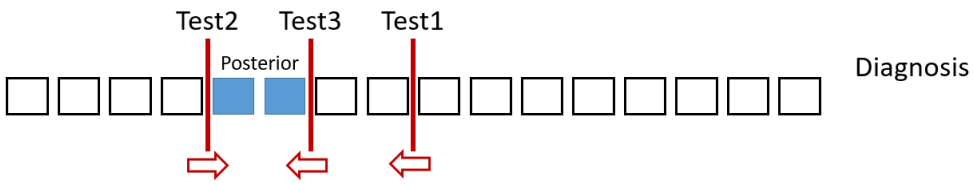


Figure 3: Belief f^2 after two signals.

thus after n binary signals the uncertainty is over $D/2^n$ states.

The choice is: how many mental signals n^M and how many external signals n^E, where

$$n = n^M + n^E.$$

Cost of information The physician faces a fixed marginal cost μ^E for each binary external signal. However, marginal cost of thinking depends on how much mental load she already incurred. The load includes mental effort on this as well as on prior patients. Let the marginal cost of mental signals $(L + n^M)$, where L is prior cognitive load and n^M represents the increasing marginal cost of thinking more about the current patient. The total cost of information then is qual to:[1]

$$C(n^M, n^E) = n^E \mu^E + n^M \Big(2L + n^M + 1\Big)/2. \tag{2}$$

Action and objective After receiving all the signals, the physician chooses treatment y to minimize the expectation of loss $|y - d|$ in the objective (1). Given her posterior belief f^n, she sets optimally $y = E[d]$. The expected loss is then equal to $(D/2^{n+1})$.[2]

The physician therefore solves the following optimization task to maximize the objective that includes both losses from suboptimal treatment as well the costs of mental and external information.

$$\max_{n^M, n^E \in \mathbb{N}_0} \quad -D/2^{n^M + n^E + 1} - C(n^M, n^E). \tag{3}$$

Summary of timing 1. Patient of an unknown state d checks in,

[1]The total cost of mental signals is a sum of the marginal costs $(L + n^M)$ for $n^m \in \{1, ..., n^M\}$.

[2]This is because width of the posterior is $D/2^n - 1$. Note that y is set at the posterior expectation. Thus, the expected distance $|y - d|$ of state d from the posterior expectation is expectation over $\{\frac{1}{2}, \frac{3}{2}, ..., \frac{D/2^n - 1}{2}\}$.

2. physician has prior knowledge f^0 about d,

3. physician chooses how much to think about the patient, n^M,

4. physician chooses how many diagnostic tests to order, n^E,

5. physician's knowledge is refined to f^n,

6. physician makes treatment decision $y = E[d|f^n]$,

7. diagnosis d is revealed.

2.1 Solution and Implications

Since the first term in the objective (3) depends on the total number of signals only, while only marginal cost of mental signals increases in the second term, physician acquires mental signals only.

Lemma 1 *There exists a pair $n^M, n^E \in \mathbb{N}_0$ that maximizes* (3).

Proof: The objective is concave and decreasing in both n^M and n^E for sufficiently large n^M or n^E. The set of possible maxima is thus finite, which implies existence.

Proposition 1 ***(Thinking vs. doing)*** *If prior load L increases, then the physician chooses to think less, i.e., n^M decreases, and orders more diagnostic tests, n^E increases.* [3]

Proof: Number of mental signals n^M is that largest $n > 0$ such that that $n < \mu^E - L - 1/2$, i.e., the physician uses mental signals as long as their marginal cost is lower than that of the external signals. The resulting n^M is clearly decreasing in L.[4] For the chosen n^M the objective (3) then takes a form $(-\alpha 2^{-n^E} - \mu^E n^E)$, where α is decreasing in n^M. Lower α pushes the objective higher, but more so for higher n^E, which increases the optimal n^E. [5]

Let us introduce a term *commonality* of a test. Higher commonality means that the test is used for more types of diagnosis d, i.e., it is more common. For instance, in Figure 3 Test1 is more common than Test2, which is more common than Test3. While Test1 us used at the beginning

[3] To be able to discuss comparative statics results despite the discrete nature of the omptimization problem, we assume tie-breakers such that both n^M and n^E are THE LOWER ONES WHENEVER ....

[4] Here the strict inequality for n uses one of the tiebreakers.

[5] TIEBREAKER

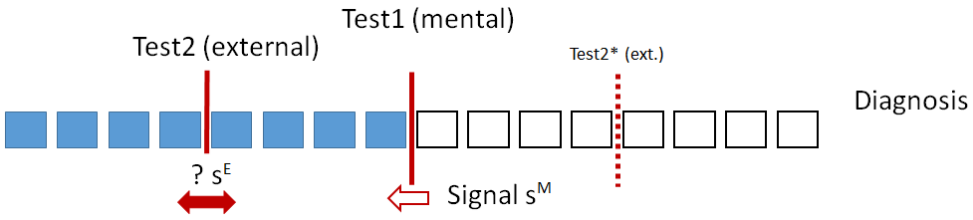


Figure 4: Belief f^1 when Test2 is the first external test ordered.

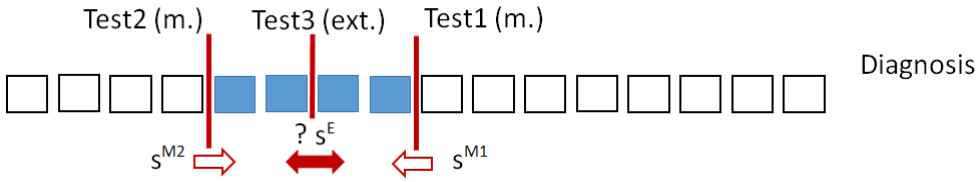


Figure 5: Belief f^2 when Test3 is the first external test ordered.

of information acquisition process and thus from all diagnoses, Test3 is applied only if the result of Test1 is "lower" and result of Test2 is "higher", which is only for the second quartile of all diagnoses. The more common tests lead to less precise posteriors are also applied more often.

The following are immediate implications of Proposition 1.

Corollary 1 ***(Uncertainty)*** *If prior load L increases, then uncertainty f^{n^M} at a moment of the first diagnostic order increases and typicality of the first diagnostic order increases, too.*

Corollary 2 ***(Commonality)*** *If prior load L increases, then commonality of the first external diagnostic test increases.*

2.2 Inference and Relation to Data

Elements of the model relate to our data very closely. Most immediate observables are state d, which is the patient's diagnosis, and cognitive load L, which we described below.

What about beliefs f? We assume that physicians have rational expectations, and thus prior f^0 coincides with the overall distribution of d. An important element for our analysis is f^{n^M}, belief after mental signals are received and when first external signal is being ordered.

We can infer average belief conditional on a particular action the physician takes. For instance, see Figure 4. There is distribution of diagnosis d if the first external test is Test2, after mental

signals pointed to lower half of d. Each test is a particular action that we observe in data, and here we identify it with a particular type of signal, i.e., a particular partition.

If the physician is rational, then the average belief at the moment of ordering the test is the distribution of diagnosis conditional on that the order was made. If on the other hand, the mental signal pointed to the upper half of diagnosis, then Test2* would be ordered instead, and the inferred belief would be different, too.

In Figure 5, there is a particular realization of beliefs after two mental signals. The external Test3 reveals beliefs at that moment. Type of diagnostic test reveals beliefs.

2.3 Discussion of Assumptions

3 Institutional Setting and Data

We study physician decisions for the universe of patients over age 18 who arrived at the University of California, San Francisco (UCSF) emergency department (ED) between January 2017 and December 2019. Our dataset comprises 71,818 visits to the ED, which we refer to as *encounters*.[6] Table 3 provides summary statistics on the demographics, health status, and chief complaints of the arriving patients. The median patient spends slightly over 5 hours in the ED. 31 percent of patients arrive via emergency vehicle, and 25 percent are admitted from the ED to inpatient status. The most commonly presented chief complaints – shown in more detail in Appendix Table A.6 – are abdominal pain (10.6% of all encounters), chest pain (6.7%), and shortness of breath (5.9%). Information on the 536 physicians is presented in Table 3. There are slightly more female than male physicians, and residents constitute the majority.

[6] ...% of encounters involve patients with multiple visits during our observation period, yielding ... unique *patients* in our sample unique *patients* in our sample. For simplicity, we refer to encounters and patients interchangeably.

Table 1: Patient demographics and health status

	Min	Mean	Median	Max
Demographics				
Share Male		0.506		
Share Asian		0.179		
Share Black		0.151		
Share White		0.459		
Age	18	51.318	52	89
Health status				
Share admitted within last 60 days		0.445		
Share ED visit within last 60 days		0.216		
Share upgraded to inpatient		0.251		
Share readmitted within next 30 days		0.041		
Share arrived in emergency vehicle		0.307		
Charlson comorbidty index	0.000	0.710	0.000	16.000
ED Length of Stay (Hours)	0.017	7.244	5.117	125.100
Inpatient Length of Stay (Days)	0.000	5.860	3.875	25.000

N = 71,818

Notes: Observations are at the *encounter*-level.

Table 2: Physician demographics

	Share
Gender	
Female	0.524
Type	
Intern/Resident Years 1 And 2	0.476
Resident Years 3 Plus	0.231
Attending Physicians	0.229
Physician Assistant	0.019
Nurse Practitioner	0.025

N = 536

The Epic Audit Log records every interaction of physicians with the Electronic Medical Record (EMR), allowing us to observe the timing and nature of all physician actions, including placing orders (medications, lab tests, etc.), note taking, and inpatient admission decisions, all with precise timestamps. Orders are placed in batches; therefore, we construct a panel at the *encounter-action* level, where each observation corresponds to an interaction between a physician and a patient. We distinguish between two main types of actions: placing orders and taking notes.

3.1 Assignment of Patients to Physicians

A key assumption of our empirical analysis is quasi-random assignment of patients to physicians. This is plausible in our setting given that the typical pattern is for patients to be assigned sequentially to available physicians. Patient's assignment will be plausibly quasi-random provided we control for the time factors that determine which physicians are working at the time of each physician-patient-interaction.

Physicians also see the chief complaint and the acuity code (1 through 5) of the patients currently waiting to see a physician (see Appendix Figure A.1). To rule out selection of physicians

to patients based on specialty (e.g., some physicians primarily treating patients with certain chief complaints), we further condition our analysis on chief complaint and acuity code. We implement this conditioning by controlling for a vector $\mathbf{T}_i$ containing hour-of-day, day-of-week, chief complaint, and acuity code indicators. A causal interpretation of our results thus requires that the additively functional form for the controls is sufficient for the conditional independence assumption to hold.

To provide evidence on the plausibility of quasi-random assignment, we look at the extent to which our vector of observable patient characteristics is balance across radiologist conditional on the minimal controls $\mathbf{T}_i$. We first define a leave-out measure of the propensity of each physician to order diagnostic tests (so the leave-out average of our main outcome variable of interest):

$$Z_i = \frac{1}{N_{j(i)} - N_i} \sum_{i' \neq i} \mathbf{1}[i' \in I_{j(i)}] \sum_n d_{i'n} \tag{4}$$

where $I_j(i)$ is the set of patients assigned to physician j, $N_{j(i)}$ is the total number of patient-interactions physician j has, and N_i is the number of interactions patient i has with physician j.

Figure 6 presents the results. It shows the individual coefficient estimates from regressions of the total number of diagnostic orders placed for the encounter (Panel A), the leave-out measure of the physician's cognitive load (Panel B), and the leave-out measure of the physician's propensity to place diagnostic orders Z_i (Panel C), on the elements of $\mathbf{X}_i$, controlling for $\mathbf{T}_i$. All outcome variables and variables in $\mathbf{X}_i$ are standardized to have mean zero and standard deviation equal to 1. At the bottom of each panel we report F-statistics for the null hypothesis that all coefficients on the elements of X_i are equal to zero and the R^2 of the vector X_i. Although $\mathbf{X}_i$ is highly predictive of the total number of diagnostic tests placed for the patient, it has far less predictive power for the measure of cognitive load or Z_i. There is further no obvious relationship between their respective coefficients in the regressions of the total number of diagnostic orders, the physician's cognitive load, and the physician's average propensity to place diagnostic tests. Overall, we see statistically significant but economically small and arguably negligible imbalance in patient characteristics. This is corroborated by small F-statistics (< 10) and small R^2 (< 0.003) in panels (B) and (C), while

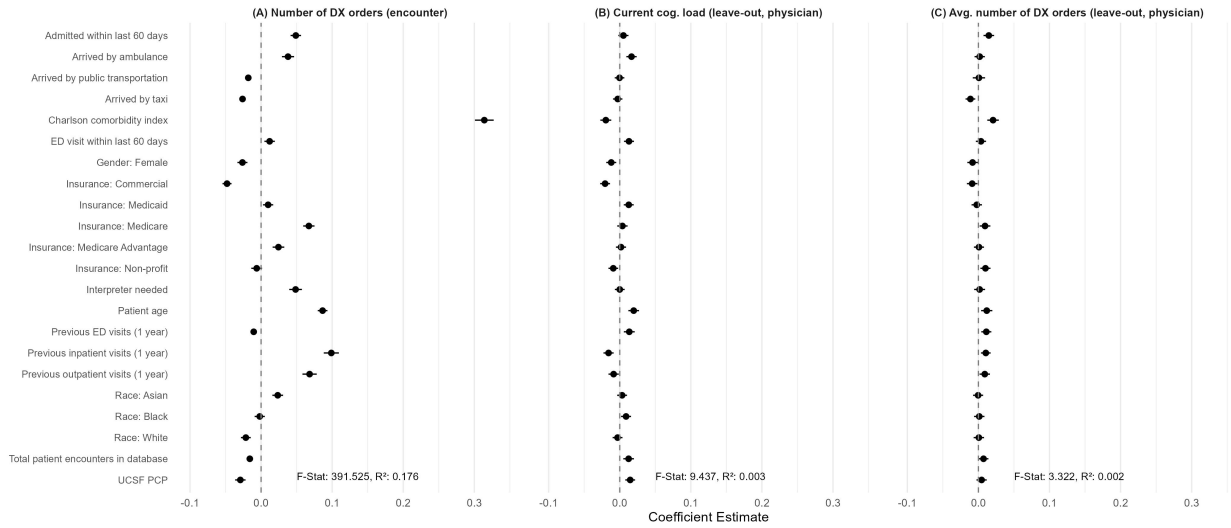


Figure 6: Covariate Balance

Notes: This figure shows coefficients and 95% confidence intervals from regressions of the total number of diagnostic orders placed for the encounter (column A), the assigned physician's leave-out current cognitive load (column B), and the assigned physician's leave-out average number of diagnostic orders placed across all encounters (column C), on covariates X_i, controlling for hour-of-day, day-of-week, hours-since-shift-start, hours-until-shift-end, acuity score, and chief complaint fixed effects. All outcome variables and covariates are standardized so that their means are equal to 0 and standard deviations are equal to 1. Each columns also prints the F-statistic. At the bottom of each panel, we report the F-statistic from the joint F-test of all covariates and the R^2 from the regression of the respective outcome variables on all covariates.

these are much larger in the regression for the number of diagnostic orders.

3.2 Measuring Expected Workload Intensity for each Encounter

For each encounter, we construct a measure of the expected workload intensity c_i based on number of orders, log note edit time, and number of consultations. Each of these variables is predicted using cross-validated LASSO regression, leveraging encounter-level data that includes age, vital signs, acuity code, means of arrival, top-level (tier 1) and mid-level (tier 2) classifications of chief complaint, Charlson comorbidity index, interpreter need, and recent ED or admission history within the past 60 days.

Appendix Tables A.1-A.3 show the LASSO coefficients of the 25 most important predictors. The most important predictors for the number of orders placed for a given encounter are non-urgent and less-urgent acuity assessments which both enter negatively. Chief complaints related to the endocrine system positively predict the number of orders, likely since this tier comprises an array of relatively distinct conditions (e.g., diabetes, thyroid disorders, hormonal imbalances).

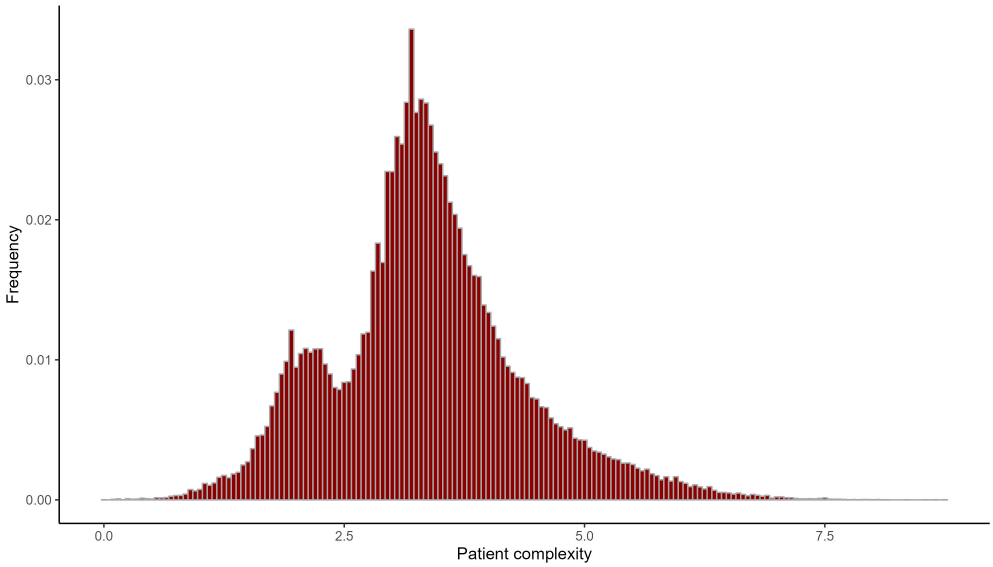


Figure 7: Distribution of patient complexity measures

This figure shows the distribution of patient complexity measures.

Non-urgent or missing acuity codes also substantially decrease the predicted note edit time for the encounter while urgent or emergent acuity assessments increase the predicted time spent editing notes. Chief complaints in the mental health category ("psychiatric" or "suicidal") lead to more consultations. Across all three variables, the LASSO coefficients always agree in their sign and are mostly extremely similar in magnitude.

We then combine the three predictions into a composite measure by summing the standardized predicted values. This composite score is further normalized to a z-score. Finally, we apply a min-shift by adding the absolute value of the minimum composite score, producing a "patient complexity" score that is both centered and non-negative across the sample.

Figure 7 shows the distribution of patient complexity measures. The distribution exhibits a right-skewed pattern, with the majority of values concentrated around the primary peak near 3.0. A smaller mode is apparent around 2.0, suggesting a substantial subset of patients who require relatively less attention.

3.3 Defining Cognitive Load

We next construct a physician's cognitive load at each action time based on these "patient complexity" scores. For physician j making an action for patient i at time t, we define the cognitive load as the sum of the encounter-level complexity measures of all distinct patients (excluding the focal patient i) for whom the physician has placed orders in the preceding 90 minutes within the same shift.[7] Since we exclude the focal patient i, this leave-out measure varies across patients i even if an action would have been conducted at the same time t. Letting $a_{i,j,t} > 0$ denote an interaction of physician j with patient i at time t, physician j's cognitive load $F_{i,j,t}$ when attending to patient i at time t is given by:

$$F_{i,j,t} = \sum_{k \neq i} c_k \tag{5}$$

where the sum is over all patients with whom the physician j has interacted at least once within $t - 90$ minutes in the current shift. Figure 8 shows the trajectory of this cognitive load measure over the course of a shift.

To account for cumulative effects from earlier in the shift, we also construct a measure of long-running fatigue. This is calculated as the cumulative sum of the encounter-level complexity measures of all distinct patients (excluding the focal patient i) for whom the physician has placed orders from the start of the shift up to 90 minutes before the current interaction:

$$G_{i,j,t} = \sum_{m \neq i} c_m \tag{6}$$

where now the sum is over all patients with whom physician j has interacted between the shift start and $t - 90$ minutes before the focal interaction.

[7] All results are robust to using a window of 60 or 120 minutes instead of 90 minutes.

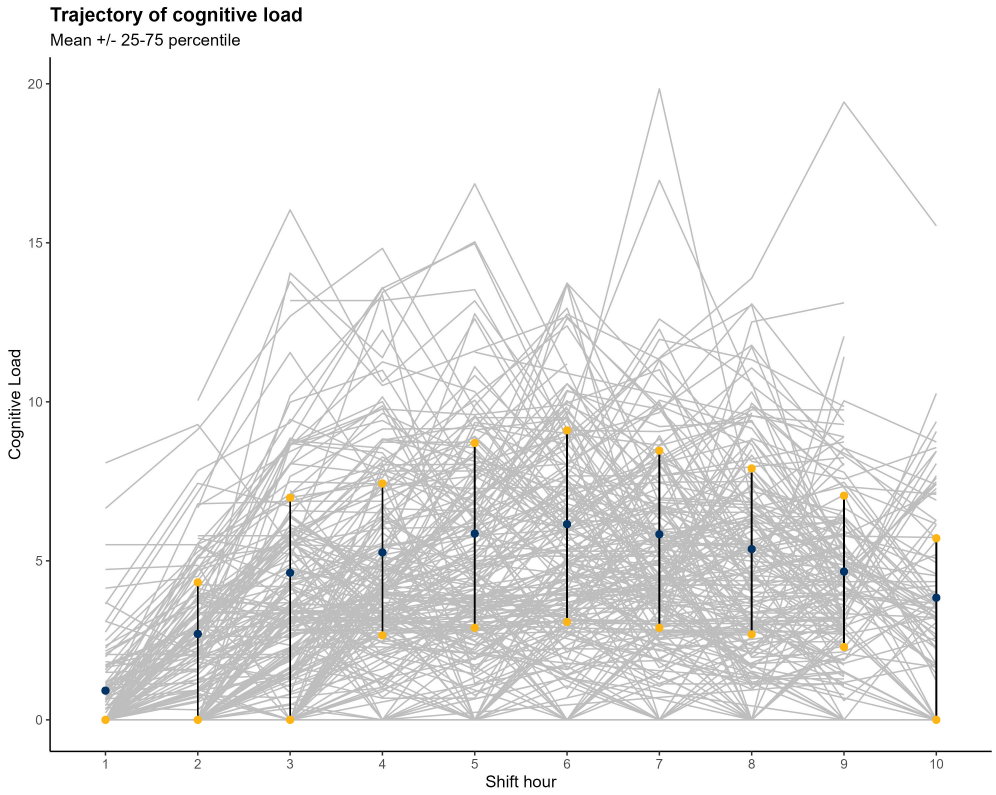


Figure 8: Trajectory of cognitive load

Notes: This figure shows the average, 25-th percentile, and 75-th percentile of the cognitive load for each shift hour across all observed shifts. Shift hour h comprises all actions conducted in the interval $(h-1, h]$; e.g., shift hour 1 includes all actions between minute 1 and 60 of the shift.

4 Empirical Approach

Exploiting the quasi-random assignment of patients to physicians, we estimate a series of regression models of the following form:

$$Y_{i,j,t} = \alpha + \beta f\left(F_{i,j,t}\right) + \gamma f\left(G_{i,j,t}\right) + \delta_t + \lambda_i + \theta_j + \varepsilon_{i,j,t} \tag{7}$$

where i indexes the patient, j indexes the physician, and t indexes time. $Y_{i,j,t}$ denotes an outcome of interest, $F_{i,j,t}$ denotes the leave-out measure of cognitive load defined in Equation (5), and $G_{i,j,t}$ represents the leave-out measure of long-run cognitive load defined in Equation (6). We consider specifications where $f(\cdot)$ is either a z-score transformation or a set of quintile indicators for the underlying variable. We include time- and shift-time-fixed effects (hour of day, day of week, hours since shift starts, hours until shift end) – captured by δ_t – and encounter-level fixed effects (chief complaint, acuity code) – captured by η_i – to ensure the causal interpretability of the coefficient of interest β (see Section 3.1). Finally, we include physician fixed effects θ_j to focus on within-physician variation.

5 Results

5.1 Resource Utilization

We begin by examining how cognitive load affects resource utilization, using the number of diagnostic orders placed as the outcome in Equation (7). Table 3 presents these results. In column (2), the coefficient on cognitive load indicates that a one standard deviation increase in cognitive load leads to 0.05 additional diagnostic orders in a batch, equivalent to a 3.6 percent increase relative to the baseline.

Since this specification conditions on the physician already placing an order batch, it captures only the *intensive margin*.[8]

[8]Our analysis ignores the positive effect of cognitive load on the *extensive* margin: a physician with a one-standard-deviation higher leave-out cognitive load places 0.117 more order batches per encounter (3.8% more relative to the 3.08 baseline) once we control for acuity code, chief complaint, day of week, median hour of day, median hours since shift

Appendix Table A.8 shows that this effect is driven largely by physicians in the top 20% of cognitive load placing more orders, whereas physicians in the bottom quintile place fewer orders; being in quintiles 2, 3, or 4 has no discernible effect on diagnostic orders.

Table 3: OLS Estimates of the Impact of Cognitive Load on Orders

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.082***	0.050***	−0.042***	−0.005
	(0.006)	(0.006)	(0.003)	(0.003)
Prior cognitive load - Control	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.39	1.39	0.82	0.82
N	256,638	256,638	256,638	256,638
R^2	0.045	0.068	0.032	0.139
Adjusted R^2	0.042	0.062	0.029	0.134

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Columns (3) and (4) report the effect of cognitive load on medication orders, providing a useful validation check. We find that an increase in cognitive load does not translate into more medication orders, suggesting that physicians are not uniformly increasing all order types, but rather selectively expanding diagnostic orders.

Columns (1)-(4) of Table A.8 document that higher cognitive load leads physicians to place more of the most common diagnostic orders (top 25) and fewer of the less common ones (outside the top 25).These results point to a shift toward a more "standardized" or "routine" set of diagnostics under higher load.

start, median hours until shift end, and mean long-run cognitive load. We focus on the intensive margin because this allows us to compute cognitive load precisely at the time of order placement rather than relying on an average (or other moments) over the course of a given encounter.

5.2 Documentation Effort

We next turn to documentation effort by examining the time a physician spends editing patient notes. Columns (5) and (6) from Table 4 reports the results from specification (7), using 181,154 note edits as the unit of observation and the natural logarithm of editing time as the outcome. On average, physicians spend approximately 1,481 seconds (about 25 minutes) on each note-taking instance. Column (6) shows that a one standard deviation increase in cognitive load reduces editing time by about 8.3 percent.[9]

Table 4: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.101***	0.065***	−0.135***	−0.020***	−0.154***	−0.083***
	(0.005)	(0.005)	(0.007)	(0.006)	(0.005)	(0.005)
Prior cognitive load - Control	X	X	X	X	X	X
Hour of day - FE	X	X	X	X	X	X
Day of week - FE	X	X	X	X	X	X
Hours since shift start - FE	X	X	X	X	X	X
Hours until shift end - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.38	0.38	1480.79	1480.79
N	256,638	256,638	256,638	256,638	181,154	181,154
R^2	0.051	0.070	0.035	0.164	0.039	0.108
Adjusted R^2	0.048	0.065	0.033	0.160	0.035	0.103

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

5.3 Diagnostic Precision

After establishing that higher cognitive load increases diagnostic orders and reduces documentation time, we now assess the impact on diagnostic precision. We construct a measure of the "infor-

[9] As for resource utilization, our analysis focuses on the *intensive margin*. On the *extensive margin*, a physician with a one-standard-deviation higher cognitive load has 0.05 more note-edit instances per encounter (+2.3% relative to the 2.17 baseline), partially offsetting the diminished effort on the intensive margin.

mativeness" of a physician's diagnostic orders with respect to possible patient diagnoses, thereby recovering the physician's implied prior over diagnoses from the diagnostic orders placed.

In the first step, we compute the empirical probability mass function (pmf) of final diagnoses for patient i receiving diagnostic order $o_{i,p}$. Since orders are used differently depending on the chief complaint, we estimate the pmf of the final diagnosis separately for each chief complaint cc_i:

$$\widehat{\Delta}_{DX}\left(o_{i,p} \mid cc_i\right) \tag{8}$$

Next, let $O_i = \{o_{i,1}, \ldots, o_{i,P}\}$ denote the set of diagnostic orders placed for encounter i. To rule out physician-learning from observing test results, we focus on orders placed within the first 30 minutes of the initial patient–physician interaction. Since we are underpowered to estimate the *joint* implied distribution over diagnoses, we construct an "aggregate" implied distribution by summing and then re-normalizing the marginal distributions of each order in the order set. We then compute the entropy of this aggregate distribution:**

$$H_i := H\left(\sum_{p=1}^{P} \widehat{\Delta}_{DX}\left(o_{i,p} \mid cc_i\right)\right) = -\sum_{d=1}^{D} p(d \mid O_i,\, cc_i) \log_2\left[p(d \mid O_i,\, cc_i)\right] \tag{9}$$

Here, $p(d)$ denotes the conditional empirical probability of diagnosis d given the chief complaint cc_i and order set O_i, i.e. implied by $\sum_{p=1}^{P} \widehat{\Delta}_{DX}\left(o_{i,p} \mid cc_i\right)$. The rationale is that a higher entropy value implies a more dispersed prior, resulting in a order set that is informative over many potential diagnoses.

In the last step, we compute the entropy over the final diagnoses associated with each chief complaint, and then subtract H_i from this value to construct a measure of entropy *reduction*. This has the advantage of an intuitive interpretation of a higher entropy reduction implying higher diagnostic precision by the physician.

$$E_i := H\left(\widehat{\Delta}\left(cc_i\right)\right) - H_i \tag{10}$$

Table 5 shows the results from the linear regression of this measure of entropy reduction E_i on cognitive load. This analysis focuses on the 48,761 encounters for which any diagnostic orders are

Table 5: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision

	Entropy reduction (cond. on CC; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.024^{***}	-0.018^{***}	-0.020^{***}	-0.014^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Number of DX orders			-0.022^{***}	-0.021^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.393^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.48	0.48	0.48	0.48
N	46,627	46,627	46,627	46,627
R^2	0.347	0.370	0.387	0.518
Adjusted R^2	0.337	0.354	0.372	0.506

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation are all orders placed within 20 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

placed within 30 minutes of the first physician-patient interaction. The coefficient in column (2) implies that a one standard deviation increase in cognitive load leads to a $\frac{0.020}{0.48} = 3.8\%$ decrease in the entropy reduction. Put differently, physicians with higher cognitive load are less precise in placing relevant diagnostic orders. Column (3) adds a control for the number of diagnostic orders to account for the mechanical effect that a larger order set is informative over more diagnoses; the coefficient of interest is basically unchanged, suggesting that the reduction in diagnostic precision among cognitively loaded physicians is not driven by these physicians simply placing more diagnostic orders. Column (4) adds the share of diagnostic orders in the 25 most common diagnostic orders to account for the mechanical effect that more frequently placed orders are associated with a wider array of diagnoses. The coefficient of interest decreases minimally. This suggests that cognitively loaded physicians are less precise in their diagnostic orders *even conditional* on the number and frequency of orders placed.

6 Counterfactual Allocation

In light of our empirical findings linking physician cognitive load to decision-making – most notably, the observed non-linear increases in diagnostic orders and alterations in inpatient admissions – we now explore an alternative patient allocation protocol that accounts for real-time cognitive load. The core idea behind this counterfactual exercise is to examine whether reassigning patients to physicians with lower current cognitive demands can mitigate the propensity for excessive or less precise ordering behavior.

Specifically, we simulate a scheduling algorithm in which arriving patients are reallocated to the physician exhibiting the lowest current cognitive load, as measured by Equation (5). In the event of ties, patient assignment is determined randomly. This counterfactual policy is motivated by our earlier results demonstrating that increased cognitive load is associated with a convex increase in order placement costs. By redistributing patients so as to balance these cognitive demands more evenly across physicians, the algorithm is designed to reduce the aggregate cognitive burden during high-intensity periods.

Appendix Figure A.2 displays the kernel density estimate of cognitive load under actual ver-

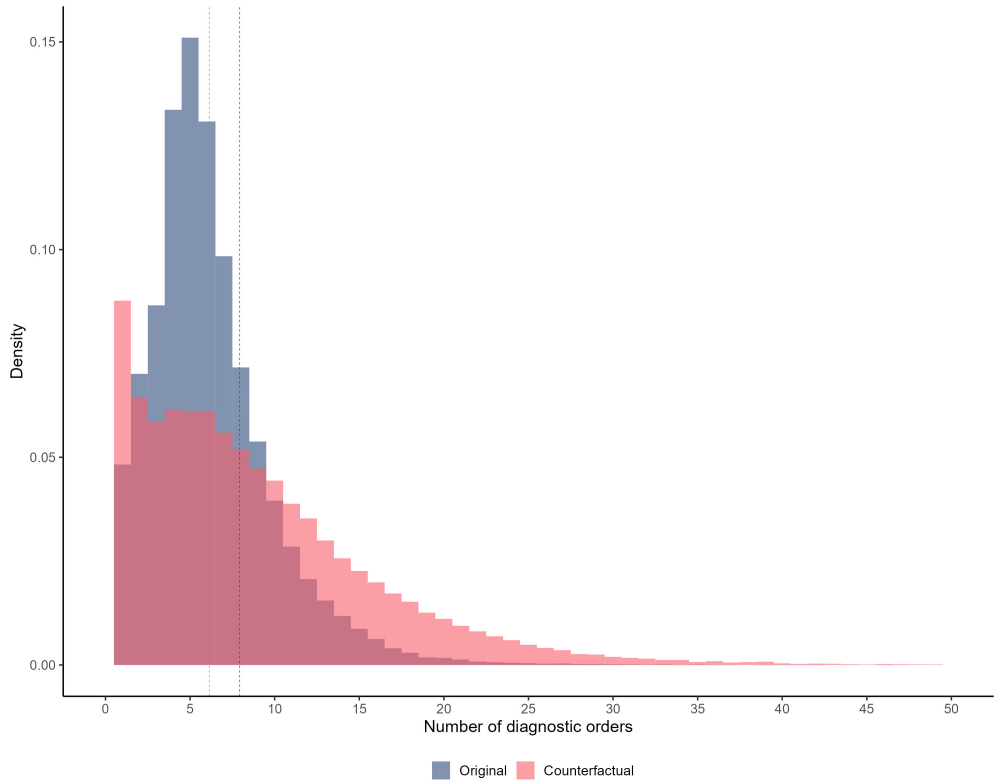


Figure 9: Distribution of number of diagnostic orders under actual and counterfactual cognitive load

This figure shows histograms of the actual number of diagnostic orders placed (blue) and the histogram under the counterfactual allocation where patients are assigned to the physician on shift with the lowest cognitive load (red). The dashed lines indicate the means under the respective regimes.

sus counterfactual patient allocations. The counterfactual distribution is shifted substantially toward lower cognitive load values, indicating that the reallocation effectively alleviates peak cognitive pressure on physicians. Complementary evidence is provided in Appendix Table **??**, which presents regression estimates allowing for non-linear effects of cognitive load on the number of diagnostic orders, respectively. These estimates serve as the basis for predicting counterfactual outcomes.

Graphical comparisons in Figure 9 illustrate the differences in the distribution of diagnostic orders under the counterfactual scenario relative to actual observations. In particular, our simulation suggests that a reallocation based on cognitive load would reduce the number of diagnostic orders by roughly 1.84 per encounter. The results indicate that a policy intervention focused on dynamic patient allocation could alleviate the cognitive burden on physicians, thereby reducing

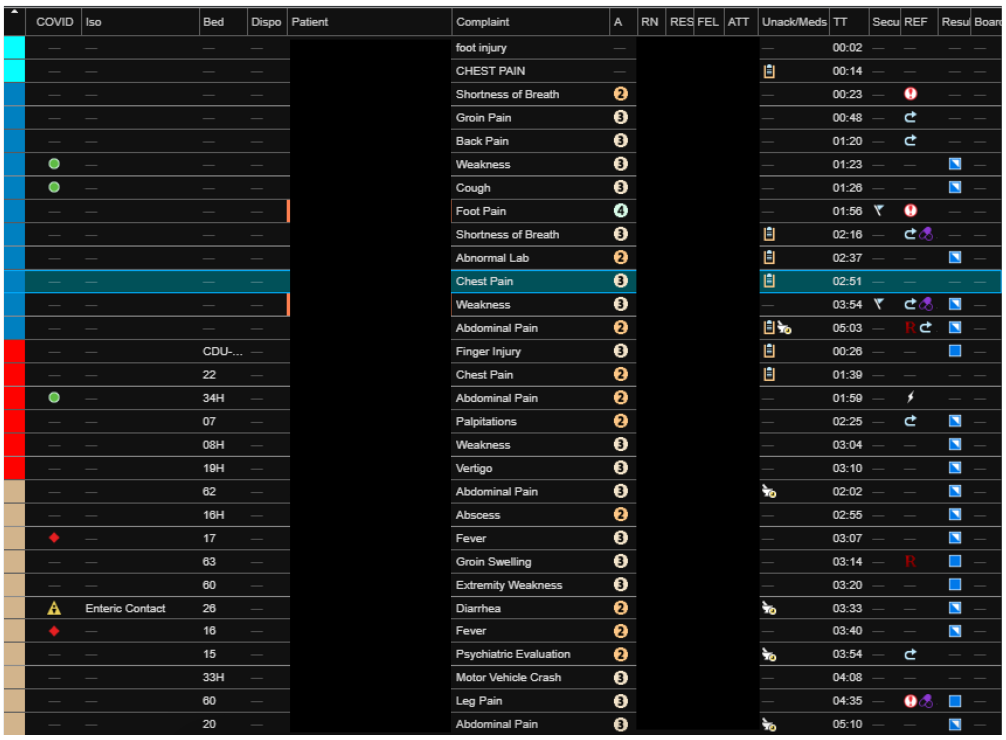
the likelihood of excessive use of diagnostic tests.

7 Discussion and Conclusion

References

A Appendix

A.1 ED Environment



	COVID	Iso	Bed	Dispo	Patient	Complaint	A	RN	RES	FEL	ATT	Unack/Meds	TT	Secu	REF	Resu	Boar
						foot injury							00:02				
						CHEST PAIN							00:14				
						Shortness of Breath	2						00:23				
						Groin Pain	3						00:48				
						Back Pain	3						01:20				
						Weakness	3						01:23				
						Cough	3						01:26				
						Foot Pain	4						01:56				
						Shortness of Breath	3						02:16				
						Abnormal Lab	2						02:37				
						Chest Pain	3						02:51				
						Weakness	3						03:54				
						Abdominal Pain	2						05:03				
			CDU-...			Finger Injury	3						00:26				
			22			Chest Pain	2						01:39				
			34H			Abdominal Pain	2						01:59				
			07			Palpitations	2						02:25				
			08H			Weakness	3						03:04				
			19H			Vertigo	3						03:10				
			62			Abdominal Pain	3						02:02				
			16H			Abscess	2						02:55				
			17			Fever	3						03:07				
			63			Groin Swelling	3						03:14				
			60			Extremity Weakness	3						03:20				
		Enteric Contact	26			Diarrhea	2						03:33				
			16			Fever	2						03:40				
			15			Psychiatric Evaluation	2						03:54				
			33H			Motor Vehicle Crash	3						04:08				
			60			Leg Pain	3						04:35				
			20			Abdominal Pain	3						05:10				

Figure A.1: Emergency Department Board

This figure shows the information screen that is available to physicians. It shows all patients who are currently admitted to the emergency department. Names of patients and physicians are removed for data privacy reasons.

A.2 Predicting Patient Complexity

Table A.1: LASSO Feature Importance: Number of all ED orders

Variable	Coefficient
Acuity code: Non-Urgent	-13.167
Chief complaint tier 1: Endocrine	11.632
Acuity code: Less Urgent	-11.248
Chief complaint tier 2: Anorectal problem	-7.943
Chief complaint tier 2: Malignancy related	-6.923
Chief complaint tier 2: Esophageal/gastric problem	-6.291
Chief complaint tier 1: GI/Abdominal	6.288
Acuity code: Missing	-5.862
Chief complaint tier 2: Flank pain	5.508
Chief complaint tier 2: Constipation	-5.361
Means of arrival: Transfer - System Arrival	-5.263
Chief complaint tier 2: Chest pain	4.835
Chief complaint tier 2: Abnormal test result	4.675
Chief complaint tier 2: Anemia	4.251
Age bin: 80+	4.078
Chief complaint tier 2: Skin/soft tissue infection	3.877
Acuity code: Urgent	-3.867
Chief complaint tier 1: Constitutional	3.738
Chief complaint tier 2: Blood disorders	3.728
Chief complaint tier 1: Neurological	3.632
Chief complaint tier 2: Liver dysfunction symptom	3.625
Means of arrival: Public Transportation	-3.559
Chief complaint tier 2: Embolic/thrombotic event	-3.485
Chief complaint tier 2: GI bleeding	-3.315
Age bin: 70-79	3.268

Table A.2: LASSO Feature Importance: Log(Number of ED note edit time)

Variable	Coefficient
Acuity code: Missing	-1.224
Means of arrival: Unknown	-0.853
Acuity code: Non-Urgent	-0.573
Chief complaint tier 2: Social problems	-0.425
Chief complaint tier 1: Drug use	-0.404
Chief complaint tier 1: New	-0.261
Acuity code: Emergent	0.252
Chief complaint tier 2: Stroke	-0.241
Acuity code: Urgent	0.192
Chief complaint tier 2: Chronic drug use problem	0.175
ED visit within last 60 days: 1	-0.173
Age bin: 30-39	-0.166
Chief complaint tier 2: Weight loss	0.161
Acuity code: Less Urgent	-0.146
Chief complaint tier 2: Blood disorders	0.134
Chief complaint tier 2: Exposure/foreign body	-0.134
Age bin: 40-49	-0.130
Age bin: 18-29	-0.126
Chief complaint tier 2: Other - iatrogenic	-0.125
Means of arrival: Walk In	-0.123
Chief complaint tier 2: Allergies	-0.113
Chief complaint tier 2: Psychiatric problem	-0.100
Means of arrival: Other	-0.099
Chief complaint tier 2: Prescriptions/referrals	-0.089
Chief complaint tier 1: Psychiatric	-0.082

Table A.3: LASSO Feature Importance: Number of Consults

Variable	Coefficient
Chief complaint tier 1: Psychiatric	0.464
Chief complaint tier 2: Stroke	0.347
Means of arrival: Police	0.323
Acuity code: Non-Urgent	-0.273
Chief complaint tier 2: Edema	-0.265
Chief complaint tier 2: Eye problem	0.239
Chief complaint tier 2: Suicidal	0.225
Acuity code: Less Urgent	-0.222
Chief complaint tier 1: Endocrine	-0.216
Chief complaint tier 2: Diabetic problem	0.207
Chief complaint tier 2: Weight loss	0.196
Acuity code: Missing	-0.154
Chief complaint tier 2: Allergies	-0.146
Chief complaint tier 2: Embolic/thrombotic event	-0.144
Chief complaint tier 2: Skin/soft tissue infection	0.142
Chief complaint tier 2: LOC	-0.142
Chief complaint tier 2: Blood disorders	-0.137
Chief complaint tier 2: GI bleeding	0.135
Chief complaint tier 2: Cardiovascular complaint	-0.128
Chief complaint tier 2: Liver dysfunction symptom	-0.128
Chief complaint tier 2: Surgical problem	-0.121
Chief complaint tier 2: Chest pain	-0.119
Chief complaint tier 2: Altered Mental Status	0.117
Chief complaint tier 2: Skin problem	0.111
Chief complaint tier 2: Esophageal/gastric problem	0.109

A.3 Additional Summary Statistics

Table A.4: Most frequently placed diagnostic orders

	Order	Share
1	Complete Blood Count With Differential	0.0925
2	Basic Metabolic Panel (Na, K, Cl, Co2, Bun, Cr, Glu, Ca)	0.0594
3	Venous Blood Gas W/Lactate	0.0467
4	Troponin I	0.0416
5	Comprehensive Metabolic Panel (Bmp, Ast, Alt, T.bili, Alkp, Tp Alb)	0.0356
6	Urinalysis With Reflex To Culture	0.0350
7	Urine Culture From Screening Urinalysis	0.0350
8	Peripheral Blood Culture	0.0324
9	Prothrombin Time	0.0299
10	Xr Chest 2 Views Pa And Lateral	0.0293
11	Lipase	0.0250
12	Magnesium, Serum / Plasma	0.0227
13	Bilirubin, Total	0.0208
14	Alanine Transaminase	0.0208
15	Aspartate Transaminase	0.0208
16	Alkaline Phosphatase	0.0207
17	Xr Chest 1 View Ap	0.0164
18	B-Type Natriuretic Peptide	0.0146
19	Poct Urine Pregnancy, ¿= 18 Years	0.0141
20	Ct Brain Without Contrast	0.0140
21	Poct Glucose	0.0137
22	Phosphorus, Serum / Plasma	0.0131
23	Ct Abdomen/Pelvis With Contrast	0.0113
24	Drugs Of Abuse Screen, Rapid	0.0108
25	Type And Screen	0.0105

Table A.5: Most frequently placed medication orders

	Order	Share
1	Sodium Chloride 0.9 % Iv Bolus	0.0850
2	Sodium Chloride 0.9 % (Flush) Injection Syringe	0.0561
3	Ondansetron Hcl (Pf) 4 Mg/2 Ml Injection Solution	0.0550
4	Acetaminophen 500 Mg Tablet	0.0395
5	Acetaminophen 325 Mg Tablet	0.0308
6	Ondansetron Hcl 4 Mg Tablet	0.0205
7	Ibuprofen 600 Mg Tablet	0.0190
8	Fentanyl (Pf) 50 Mcg/Ml Injection Solution	0.0165
9	Oxycodone 5 Mg Tablet	0.0150
10	Electrolyte-A Iv Bolus	0.0149
11	Lorazepam 2 Mg/Ml Injection Solution	0.0132
12	Lorazepam 1 Mg Tablet	0.0126
13	Potassium Chloride 10 Meq/100ml In Sterile Water Intravenous Piggyback	0.0122
14	Morphine 4 Mg/Ml Intravenous Solution	0.0110
15	Hydromorphone (Pf) 0.5 Mg/0.5 Ml Injection Syringe	0.0104
16	Ipratropium Bromide 0.02 % Solution For Inhalation	0.0101
17	Ketorolac 15 Mg/Ml Injection Solution	0.0099
18	Potassium Chloride Er 20 Meq Tablet,Extended Release(Part/Cryst)	0.0094
19	Vancomycin 1 G Ivpb (Vial-Mate)	0.0090
20	Acetaminophen 1,000 Mg/100 Ml (10 Mg/Ml) Intravenous Solution	0.0089
21	Hydrocodone 5 Mg-Acetaminophen 325 Mg Tablet	0.0089
22	Sodium Chloride 0.9 % Intravenous Solution	0.0084
23	Ceftriaxone 1 G Iv 50 Ml (Mini-Bag Plus)	0.0082
24	Lidocaine (Pf) 10 Mg/Ml (1 %) Injection Solution	0.0082
25	Diphenhydramine 50 Mg/Ml Injection Solution	0.0080

Table A.6: Most common chief complaints

	Chief complaint	Share
1	Abdominal Pain	0.1062
2	Chest Pain	0.0675
3	Shortness Of Breath	0.0591
4	Fall	0.0302
5	Altered Mental Status	0.0289
6	Headache	0.0261
7	Dizziness	0.0250
8	Weakness	0.0244
9	Fever	0.0241
10	Psychiatric Evaluation	0.0238
11	Back Pain	0.0236
12	Emesis	0.0181
13	Leg Pain	0.0167
14	Other	0.0165
15	Syncope	0.0158
16	Flank Pain	0.0151
17	Seizures	0.0145
18	Cough	0.0140
19	Alcohol Intoxication	0.0130
20	Suicidal	0.0128
21	Knee Pain	0.0093
22	Palpitations	0.0091
23	Foot Pain	0.0089
24	Abnormal Lab	0.0085
25	Leg Swelling	0.0081

N = 71,818

Table A.7: Most common primary diagnoses

	Diagnosis	Share
1	Chest Pain, Unspecified	0.0356
2	Unspecified Abdominal Pain	0.0301
3	Sepsis, Unspecified Organism (Cms Code)	0.0255
4	Other Chest Pain	0.0190
5	Syncope And Collapse	0.0144
6	Dizziness And Giddiness	0.0139
7	Headache	0.0119
8	Urinary Tract Infection, Site Not Specified	0.0104
9	Epigastric Pain	0.0098
10	Pneumonia, Unspecified Organism	0.0096
11	Unspecified Convulsions (Cms Code)	0.0076
12	Palpitations	0.0073
13	Shortness Of Breath	0.0073
14	Unspecified Fall, Initial Encounter	0.0071
15	Low Back Pain	0.0071
16	Chronic Obstructive Pulmonary Disease With (Acute) Exacerbation	0.0070
17	Calculus Of Kidney	0.0068
18	Nausea With Vomiting, Unspecified	0.0068
19	Fever, Unspecified	0.0067
20	Abnormal Electrocardiogram (Ecg) (Ekg)	0.0059
21	Hypertensive Heart And Chronic Kidney Disease With Heart Failure	0.0056
22	Altered Mental Status, Unspecified	0.0055
23	Fall On Same Level From Slipping, Tripping And Stumbling	0.0053
24	Major Depressive Disorder, Single Episode, Unspecified	0.0053
25	Right Lower Quadrant Pain	0.0053

N = 71,818

A.4 Alternative Regression Specifications

Table A.8: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Q5 cog. load	0.165***	0.092***	−0.122***	−0.006
	(0.015)	(0.015)	(0.010)	(0.008)
Q4 cog. load	0.051***	0.018	−0.066***	−0.002
	(0.014)	(0.014)	(0.010)	(0.009)
Q2 cog. load	−0.018	−0.007	−0.059***	−0.024**
	(0.014)	(0.015)	(0.010)	(0.009)
Q1 cog. load	−0.016	−0.033**	0.019*	0.039***
	(0.015)	(0.016)	(0.011)	(0.011)
Prior cognitive load - Control	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.39	1.39	0.82	0.82
N	256,638	256,638	256,638	256,638
R^2	0.045	0.068	0.033	0.139
Adjusted R^2	0.042	0.062	0.030	0.134

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on quintiles of the leave-out measure of cognitive load. The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

A.5 Counterfactual patient allocation

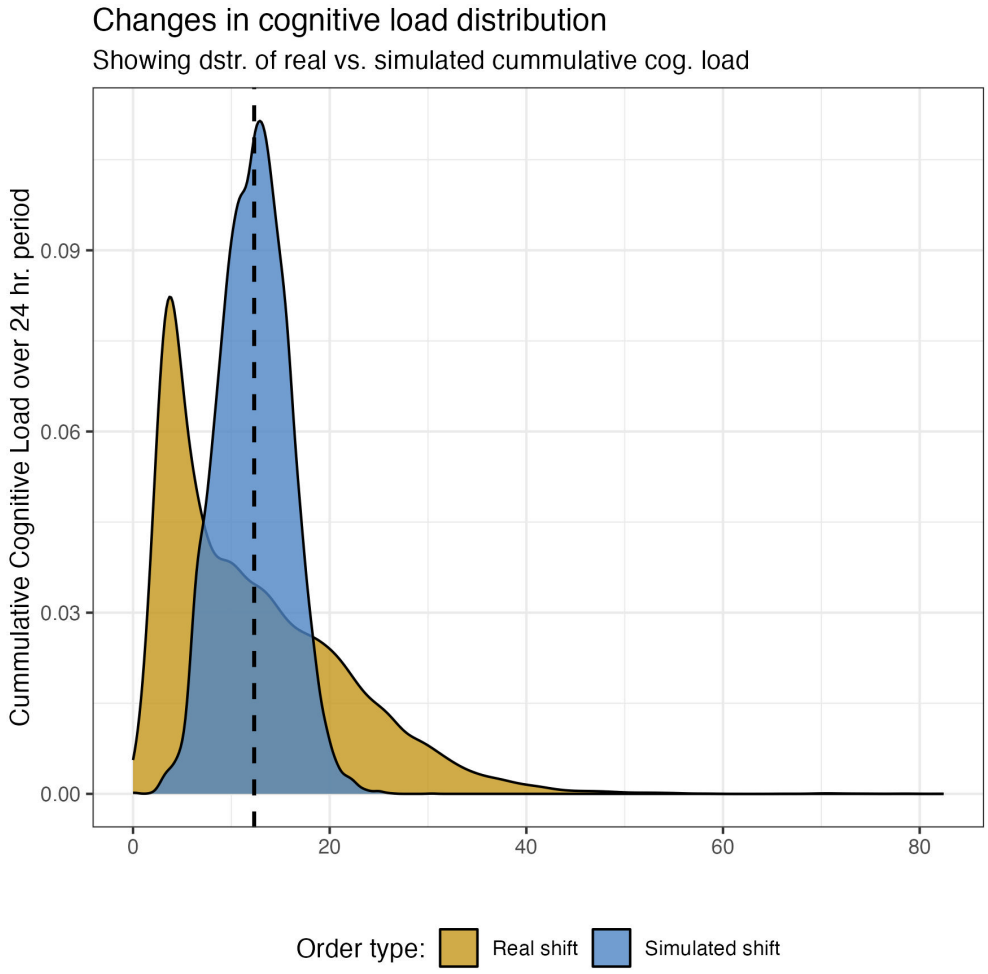


Figure A.2: Actual vs. counterfactual cognitive load

This figure shows kernel density estimates of the actual distribution of cognitive load ("real shift", in blue), and the counterfactual distribution of cognitive load ("simulated shift"). The counterfactual allocation assigns patients to the physician on shift who has lowest cognitive load upon arrival of the patient in the emergency department.

Table A.9: Number of diagnostic orders on cognitive load (cross-section)

	Number of diagnostic orders
Max cog. load decile 2	−0.014
	(0.013)
Max cog. load decile 3	−0.011
	(0.014)
Max cog. load decile 4	0.054***
	(0.014)
Max cog. load decile 5	0.079***
	(0.014)
Max cog. load decile 6	0.107***
	(0.015)
Max cog. load decile 7	0.157***
	(0.015)
Max cog. load decile 8	0.171***
	(0.015)
Max cog. load decile 9	0.210***
	(0.016)
Max cog. load decile 10	0.240***
	(0.017)
log(Patient Complexity)	1.278***
	(0.009)
Max prior cog. load - control	X
Physician - FE	X
Max shift hour - FE	X
N	71,796
R^2	0.331
Adjusted R^2	0.324

Notes: *** 1 percent level; ** 5 percent level; * 10 percent level