Thinking versus Doing: Cognitive capacity, decision making and medical diagnosis

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

November 22, 2025

Abstract

We study how situational fluctuations in cognitive capacity shape behavior in high-stakes, real-time decision-making. Drawing on recent advances in behavioral economics that revolve around inattention, cognition and complexity, we show that cognitive load influences how physicians in emergency departments allocate mental effort, attention, and make diagnostic and treatment decisions. We use quasi-random variation in patient-physician pairings, along with granular electronic medical record and audit-log data from many clinical interactions, to show that, under higher cognitive load, physicians substitute mental deliberation with more numerous but less precise diagnostic actions. Specifically, we document that higher load (i) increases the total number of orders of diagnostic tests (ii) reduces the use of more precise and less common tests (iii) increases the use of common tests and (iv) reduces the use of uncommong tests and (v) increases uncertainty in diagnostic beliefs. We show that cognitive load impacts downstream inpatient admission from the emergency department: a physician in the highest cognitive load decile increases admissions by 28% relative to the same physician in the lowest cognitive load decile, for the exact same kind of patient. We also explore implications for policy, including how patient-physician matching might be improved by accounting for cognitive load profiles. These results offer novel field-based evidence on the dynamics of attention and belief formation, and shed light on how cognitive constraints shape diagnostic behavior in complex, real-world environments.

Keywords: CHANGE Cognitive Load, Physician Decision-Making, Health Care
JEL Classification: I11, D83, D91

*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 and CEPR. Email: fmatejka@gmail.com. We thank David Chan, David Silver, and XXXX other discussants for helpful comments. We thank seminar participants at the 2025 AEA meetings, Amazon, Berkeley, BYU, Duke, CESifo Summer Institute, Georgetown, Harvard, Harvard Business School, Imperial College London, LSE, MIT, Northwestern, Stanford, University of Chicago, University of Zurich, and University of Tokyo for their helpful feedback. We thank Julia Adler Milstein for her collaboration in accessing the data.

1 Introduction

All decision-making requires mental effort. Agents in the changing and complex economy need to filter out new information from numerous sources (e.g., media, customers, coworkers, data) and integrate it with their existing knowledge to form their behaviors (e.g., on the job decisions, personal choices, tacit learning). These evaluations and decisions often occur under conditions with significant bandwidth constraints, as well meaningful levels of institutional complexity. Recent advances, mostly theoretical and experimental, highlight the importance of inattention, cognitive capacity, and complexity in shaping economic decisions (e.g., Maćkowiak et al. (2023), Gabaix (2019), Bordalo et al. (2025), Caplin and Dean (2015), Oprea (2024) and Enke and Graeber (2023)). In fact, integrating such cognitive factors into economic models can potentially unify seemingly disparate behavioral phenomena under a common framework.

A key gap in the literature, however, is the lack of empirical exploration of high-frequency decision-making in real-world settings. Such empirical investigations are not only valuable for their applied relevance, but also help to inform and refine existing models of cognitive resource allocation. In this paper, we begin to fill this gap by exploring a practical, high-stakes, high-frequency setting: the diagnostic choices of physicians in emergency departments (EDs). In this setting, physicians must obtain information from multiple sources (e.g., patient history, physical examinations, conversations, imaging, and lab tests), synthesize these data into diagnostic hypotheses, and ultimately arrive at a treatment plan.

These decisions are cognitively demanding and made under time pressure, uncertainty, and institutional complexity.[1] Importantly, physicians treat hundreds or thousands of patients over time, many of whom present with similar symptoms, offering an opportunity to study decision variation within-physician over repeated, comparable choice settings. Since cognitive capacity is difficult to measure in administrative data, much of the prior literature on cognitive load has focused 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") is often accompanied 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.

We show that fluctuations in cognitive resources meaningfully affect physicians' behavior over time, even for the same physician facing similar decisions. Variation in cognitive effort influences key aspects of physicians' clinical practice, including the number and precision of diagnostic test orders, as well as costly inpatient hospital admission decisions. Beyond providing insight into the

[1] Since the seminal work of Arrow (1963) a large literature has developed to study aspects of physician decision-making (see, e.g., McGuire (2000) and Chandra et al. (2011) for overviews). This work is largely grounded in standard economic models of incentives, information, and market structure with little to no focus on within-physician variation in behaviors and, especially, the cognitive constraints of physicians.

diagnostic process in the ED, our findings yield broader behavioral conclusions: the cost of mental effort for experts is not fixed over time, and care approaches change flexibly, even over short time horizons. Moreover, physician behavior exhibits attenuation—a central feature of inattention-based models—where diagnostic responses under high cognitive load are systematically muted relative to those under lower load.

Specifically, we (i) develop a conceptual framework that formalizes the role of cognitive resources in diagnostic decision-making and (ii) leverage new granular data from the Emergency Department of the University of California at San Francisco (UCSF) to empirically study the implications of cognitive capacity on care. Our conceptual framework models an individual physician's pursuit of a patient's diagnosis. The physician has a level of cognitive load throughout the patient encounter that depends on the other patients she is treating or has treated. The patient arrives with a chief medical complaint and some other characteristics, from which the physician forms a prior about potential diagnoses. In sequence, the physician (i) exerts mental effort to refine her diagnostic beliefs and (ii) orders external tests to further refine those beliefs. The marginal costs of mental effort by the physician are increasing in both cognitive load and current mental effort. The physician exerts mental effort and orders external tests to the point where the benefits of further refining beliefs about the true diagnosis are outweighed by the costs of additional mental effort / external tests.

The model has several predictions that we test in our empirical analysis. First, the model predicts that a higher level of cognitive load (reduced cognitive capacity) leads to the physician exerting less effort to acquire mental signals and placing more external orders instead. This is because the marginal cost of external orders does not vary with cognitive load. Second, an immediate implication of the first finding is that higher cognitive load also leads to more uncertainty by the physician when starting to choose what external test to order. Ordered external tests will therefore be more common and less targeted towards the final diagnosis in case cognitive load is higher. Third, physicians exhibit behavioral attenuation whereby diagnostic responses under high cognitive load are systematically muted relative to those under lower load. For each of these model predictions, we develop and test an empirical analog in our data.

Fourth, our framework illustrates a revealed preference technique—based on Caplin and Dean (2015)—to infer physician beliefs from choice data: under rational expectations, the distribution of final diagnoses conditional on a given test order reveals the physician's belief at the time of ordering. This technique works as long as the choice set is sufficiently rich and physicians can tailor behavior to their beliefs. The reduction in belief uncertainty, relative to the prior, captures the information acquired, which the model shows is greater when cognitive capacity is higher. We leverage our detailed data to implement this technique and empirically study information acquisition under cognitive load.

The data include rich information on both patients and providers including (i) patient charac-

teristics (ii) medical care (e.g., orders, procedures, triage, diagnoses) (iii) physician characteristics (iv) note actions where physicians take clinical notes on patient encounters and (v) an audit log of physician click-by-click actions when placing medical orders / notes or reviewing the outcomes of prior orders / notes.[2] These data include all patients visiting the Emergency Department over a two-year period (mid-2017 to mid-2019) and include to-the-second timestamps for all relevant actions. Crucially, the detailed data allow for the construction of short-duration within-physician shift cognitive capacity measures while our environment in the ED generates quasi-random variation in those measures over time within-physician.

Our dataset has several features that make it very well suited for the exploration and application of theories of cognitive resource allocation. First, physicians repeatedly face distinct diagnostic decisions across thousands of patient encounters, allowing us to study joint statistical properties of patients' states and physicians' choices. Second, the decision space is rich, with many possible diagnoses, diagnostic orders, and treatment options. This richness is not only important for studying flexibility in allocation of cognitive resources but it is also essential to be able to address "thinking" and infer physicians' beliefs, which are revealed by the detailed choices they make while in the diagnostic process. Third, the environment involves trained professionals, enabling us to focus on cognitive limits of short-term decisions rather than on humans' long-term learning about how to approach a particular task at hand.

Empirically, we resolve two key identification issues in order to establish the impact of cognitive load on different outcomes. First, we address the issue that actual observed outcomes for a given patient are endogenous to the level of their physician's cognitive load. We develop a model of predicted patient complexity / intensity based on granular information ex ante to the patient being treated by the physician. This information includes details like (i) patient chief complaint at a granular level, (ii) patient vital signs, and (iii) the triage acuity score given to the patient by the intake nurse. We use a LASSO model that predicts, given this information for each patient, the expected (i) number of orders, (ii) amount of provider time editing notes, and (iii) number of outside consults to specialists.[3] The output of this prediction problem is a measure of expected patient complexity / intensity that does not depend on the patient's actual experienced treatment.

For a given provider, we then aggregate these patient-level predictions within-shift and across patients into granular measures of cognitive load. Our primary measure of cognitive load relates to the recent burden of complex patients a physician treats. For a given time window (e.g., 90 minutes), we assess the total complexity burden for a physician by summing up predicted patient complexity across all patients the physician is currently treating (excluding the focal patient). For a given physician and given shift, at each point in time we have a measure of contemporaneous

[2]This granular audit log data provide a uniquely granular measures that map to real-life clinical care, as described in Kannampalil and Adler-Millstein (2022). The data cover the entire time the patient is in the Emergency Department (ED) as well as the time the patient is in inpatient, if admitted to the hospital.

[3]The model residualizes out date and time fixed effects.

burden as well as lagged burden from earlier in the shift. We utilize both the current and lagged measures in our empirical analysis to reflect physician cognitive load. For each patient-physician pair, we construct these measures "leaving out" the contribution of the index patient we are studying. As described in detail in the main text, we assess robustness for these measures on a number of key dimensions including, e.g., (i) the time intervals used, (ii) how we combine the number of patients and complexity of each of the patients, and (iii) variations of the underlying LASSO prediction model.

With these measures in hand, the second identification challenge we resolve is ensuring that we have quasi-random variation in cognitive load, within and across physicians. To this end, we leverage both quasi-random patient arrival to the ED as well as quasi-random assignment of patients to physicians once they have arrived in the ED.[4] We present evidence on within-physician variation in cognitive load over the two years to illustrate the significant variation coming from quasi-random patient arrival. We present balance tests on patient characteristics to highlight the quasi-random nature of patient assignment to physicians, which we also support with qualitative evidence on the assignment process.

We set up our main analyses at the provider-patient-event level. We study the impact of provider cognitive load on key outcomes, conditioning on time of day, time of shift, day of week, acuity code and chief complaint fixed effects. In addition, we run most specifications with provider fixed effects, isolating the impact of cognitive load within-provider over time, relative to their baseline behaviors. A critical aspect of our model is the nature of time in the treatment process: we focus on current cognitive load by measuring the level of expected complexity for assigned patients in the prior 90 minutes when an action (e.g. ordering a test) is made. We include lags related to cognitive load prior to that period to control flexibly for the stock of cognitive load.[5] We allow for the effects of cognitive load to enter linearly as well as more flexible indicators for quintiles and deciles. For each observation, we compute cognitive load measures "leaving out" the contribution of the index patient to provider cognitive load.

We first test our model's prediction that higher cognitive load will lead to an increase in the number of diagnostic tests. Our main outcome variables are the number and types of tests in the order batches for each patient.[6] We find that increasing physicians' contemporaneous cognitive load by 1 standard deviation has a meaningful effect on the number of diagnostic tests in the first order batch, increasing them by 4% holding all else fixed. As a placebo, we estimate the same model for medication orders, which are more standardized and require less thinking, finding

[4]Our analysis here follows a range of recent papers that use quasi-random variation in ED patient arrival and assignment to study different questions, including, e.g., Chan (2016), Chan (2018), Mullainathan and Obermeyer (2019), and Silver (2020).

[5]We show in Appendix F.3 that our primary results are robust to alternative time windows including 60 and 120 minutes.

[6]Medical orders are typically placed in batches, where a number of orders are input simultaneously. We use this as a natural unit of observation, unpacking the characteristics of these order batches as described.

a precise null effect of cognitive load on these orders. For our specification using quintiles of physician cognitive load, we find that the top quintile of cognitive load increases the total number of diagnostic orders by 9% relative to the bottom quintile.

We investigate how the profile of diagnostic orders changes with changes in cognitive load. We find that the overall increase in diagnostic orders due to higher cognitive load masks a significant underlying shift from less common orders to more common orders. A 1 standard deviation increase in cognitive load increases the number of common orders (top 25 diagnostic orders) by 7% but decreases rare orders (outside of top 25) by 5%. This is consistent with our model where providers exert less mental effort per patient when cognitively loaded, but substitute for that to some extent with more common / standard medical tests.

Next, we turn our focus to another prediction of our model and investigate the precision of the diagnostic process under cognitive load. Following our model, we develop a novel empirical approach to study refinement of beliefs. We base our approach on statistical entropy reduction, a measure used in machine learning and information theory that quantifies the level of information gain via uncertainty reduction. We do this for a given set of diagnostic orders, assessing the extent to which those orders move us from an ex ante prior towards a final diagnosis.

We proceed in two steps. First, for each possible order and ex ante information set (chief complaint and intake acuity information) we construct the probability distribution of final diagnoses conditional on placing that order. For an order batch, we combine these distributions across the orders in the batch. Second, we compute the Shannon entropy of this implied final diagnosis distribution, which gives us a measure of how targeted the order batch is towards uncovering specific final diagnoses. Our key measure focuses on the entropy reduction of the first order batch. This measures, relative to the distribution of possible diagnoses associated with the ex ante information set (chief complaint and acuity assessment), how much more targeted is the first order batch towards particular diagnoses. The higher the entropy (lower reduction in uncertainty from the prior) the less precise physician's beliefs are and the less the physician learned about the diagnosis up to the moment of placing orders. This measure relates directly to our conceptual model, where a provider assesses a patient with thinking effort and then, conditional on that, forms beliefs and orders tests.

We find that a 1 standard deviation increase in cognitive load decreases the entropy reduction of the first order batch by 5%. This means that, as provider cognitive load increases, the first set of diagnostic tests becomes less targeted towards specific end diagnoses, implying less thinking up front and more dispersed knowledge when placing orders. In the specification with quintiles of cognitive load, being in the top quintile of cognitive load leads to a 4% decrease in entropy reduction relative to the median quintile of cognitive load. The change in entropy reduction is especially large moving from the fourth to the fifth (top) quintile of cognitive load, illustrating the non-linear relationship between our cognitive load measure and the targeting of diagnostic tests.

We investigate the impact of cognitive load on several key outcomes including (i) inpatient admission and (ii) readmission to the hospital within 30 days. To do this, we run regressions at the provider-patient encounter level, since each outcome relates to the whole course of treatment in the ED. We use the maximum point-in-time cognitive load for the physician over the entire patient encounter as our measure of cognitive load.

We find a large effect of increased cognitive load on hospital admission: under our liner model a 1 standard deviation increase in maximum cognitive load increases the chance of admission by 9%, controlling for provider, chief complaint, acuity level, day and time fixed effects. We also implement a more flexible model that allows for cognitive load deciles to enter as binary indicator variables. We find a large and near monotonic effect of cognitive load moving from low to high deciles: the admission probability of a patient is 28% higher for a physician in the highest cognitive load decile relative to one in the lowest decile. We investigate the mechanisms underlying this result and find that this large impact of cognitive load is not due to ED capacity constraints being correlated with situations of higher cognitive load. Our results on reduced diagnostic precision under higher cognitive load suggest that provider risk aversion is a plausible explanation for these admission increases under higher cognitive load. In a subsequent analysis, we find a precise zero effect for the impact of maximum cognitive load on hospital readmissions within 30 days, which is consistent with the hypothesis of increased provider risk aversion leading to more hospital admissions.

We run our primary analyses for two specific chief complaints: (i) abdominal pain (9% of our sample) and (ii) chest pain (4% of our sample). These are two of the most common chief complaints and also have meaningful degrees of freedom in diagnosis and treatment. These results corrobate our primary findings and allow us to investigate diagnostic impacts with additional contextual specificity. For abdominal patients, a 1 standard deviation increase in cognitive load increases the number of orders by 4% and does not increase medication orders. It increases the number of common orders by 9% and reduces the number of rare orders by 12%. It reduces note edit time by 9%, and decreases entropy reduction by 4%, though this last result is marginally statistically insignificant. Our results for chest pain are quite similar, highlighting the consistency of results in different specific disease categories.

We run a range of additional specifications of interest, including, e.g., (i) encounter-level regressions for all key outcomes that use maximum cognitive load over the encounter and (ii) event-level regressions that focus only on the first order batches for patients, i.e., the beginning of the diagnostic process. Our encounter-level results strongly support (and often exceed) our primary event-action-level results. In our main specification that uses deciles of maximum cognitive load during the encounter, we find a strong and monotonic relationship on the impact of cognitive load on the number of diagnostic orders. The top decile in terms of cognitive load increases the number of diagnostic orders by 42% relative to the bottom decile over the encounter, holding all

else equal. We also investigate an additional outcome, note edit time. We find that a 1 standard deviation increase in cognitive load decreases note edit time for the index patient by 7%, and that moving from the bottom to top quintile of cognitive load reduces that note edit time by 23%.[7]

We use our encounter-level estimates to perform a counterfactual calculation where we ask how a scheduling algorithm that accounts for cognitive load impacts the diagnostic tests ordered as well as the rate of hospital admissions. We first consider a simple reallocation mechanism where arriving patients are allocated to the doctor with the lowest current cognitive load. This "greedy" allocation algorithm is simple to implement and allows us to easily model changes with existing data. We hold fixed the set of physicians in the ED at a point in time as well as the number of patients assigned to each. We show that this kind of reallocation has a large impact on tails of high cognitive load. We increase the share of time spent at low cognitive load levels and have a large impact decreasing the share of time spent in potentially very high states of cognitive load. This increases order precision, decreases the number of orders, and reduces hospital admissions. We show that this algorithm achieves these results without meaningfully changing the type of physician treating patients (e.g., resident or attending).

This paper contributes to the growing theoretical and empirical literature on attention, cognitive load and complexity in decision making. Maćkowiak et al. (2023) presents an overview of rational inattention and how it relates to phenomena studied in behavioral economics, highlighting the conceptual implications for decision-making and empirical evidence in this area. Gabaix (2019) also provides a conceptual overview in the space of behavioral inattention. Enke and Graeber (2023) presents an overview of related work in this area focusing specifically on complexity and cognitive burden, highlighting advances and recent empirical work, the preponderance of which is in a lab experiment context. The vast majority of empirical work in these areas is in the context of lab experiments including, e.g., Dean and Neligh (2023) and Oprea (2024). This is because it is difficult to obtain administrative data that are granular enough to generate short-duration cognitive load measures while also finding clear factors that shift those measures over time for decision-making agents. One main methodological contribution of our paper is that we can empirically recover physicians' diagnostic beliefs and study how they vary with cognitive constraints. To do so, we incorporate the approach of "revealed posteriors" of rationally inattentive agents in Caplin and Dean (2015) —to our knowledge this is its first application in a field setting. It is also worth noting that there are literatures in other domains that use proxies for fatigue or load to study the impacts on expert behavior including, e.g., in the domain of judges (Danziger et al. (2011), Kleinberg et al. (2017)) or manufacturing workers (Kaur et al. (2025)).

While there are many papers in health economics focused on cross-sectional variation in provider behaviors (e.g., due to skill, preference, or training differences) there are few papers that study within-physician changes in behavior, and none that we are aware of that study physician cog-

[7] See Hsu et al. (2023) for a discussion of provider note edit time as mental effort.

nitive load in depth.[8] Chan (2018) and Silver (2020) study within-physician changes in behavior, and do so as we do in the ED context. Chan (2018) finds that physicians near the end of their shifts slack off by taking on fewer patients and spending more per patient while Silver (2020) finds that physicians change their behaviors in response to their peer groups and assimilate towards group production norms. Agarwal et al. (2023) perform an experiment assessing how radiologist practices change in response to information interventions, studying the interplay between human expertise and advice from artificial intelligence (AI) tools. Other notable papers studying within-physician behavior changes include Steiny Wellsjo (2025) and Chodick et al. (2025).

Several recent papers in this literature hone in on the diagnostic process in a cross-sectional sense, including, e.g., Song et al. (2010), Abaluck et al. (2016), Currie and MacLeod (2017) Mullainathan and Obermeyer (2019), and Chan et al. (2022). These papers generally illustrate meaningful cross-sectional variation in diagnostic practices and seek to unpack different underlying mechanisms. Relative to all of these prior papers, we (i) study cognitive load and fluctuations in physician decision-making capacity (ii) leverage granular administrative to both measure cognitive load and structurally uncover physician diagnostic beliefs and (iii) highlight the important implications of these micro-foundations for health care.

The remainder of the paper proceeds as follows. Section 2 presents our conceptual framework, studying the impact of cognitive load on key behaviors and outcomes. Section 3 describes our data and context. Section 4 sets up our main analysis by constructing our cognitive load measures and assessing threats to identification. Section 5 presents our main results and Section 6 concludes.

2 Conceptual Framework

In this section we first formulate a model of a physician that resolves uncertainty about a patient's diagnosis. Then, we describe the model's implications and how its elements connect to observables in data and also how the model helps us interpret our data. Finally, we discuss the model's simplifying assumptions.

2.1 Model

Physician faces a patient of an unknown state (diagnosis) $d \in \{1, ..., D\}$, where $D > 1$ is the number of possible diagnoses. Let f be physician's subjective belief about a patient's state, i. e., f is a probability distribution over the D possible diagnoses. We denote f^0 to be a prior belief, and f^n the be a belief after n signals are received. We assume that f^0 is uniform, i. e., $f^0(d) = 1/D \forall d$.

The physician chooses how to learn about state d and what treatment $y \in \{1, ..., D\}$ to prescribe

[8] See Skinner (2011) and Cutler et al. (2019) for surveys of the physician practice variations literature. See, e.g., Chandra and Staiger (2007), Finkelstein et al. (2016), and Badinski et al. (2023) for specific examples.

to maximize the following objective:

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

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 as the physician is more likely to identify the optimal treatment $y^* = d$, or at least to match treatment y more precisely to a given d. The cost of information in the second term is described below. It depends on how much cognitive effort (mental signals n^M) the physician deploys and how many diagnostic tests (external signals n^E) she uses to determine d.

Information acquisition Learning occurs when the physician acquires mental signals n^M from thinking about the patient, and then acquires external signals n^E from doing diagnostic tests. We assume that all signals have a form of a binary partition. Given a current belief $f^n = U[a, ..., b]$, the physician finds out whether the state is in a lower or an upper half of the states, i. e., f^{n+1} is either $U[a, ..., \frac{a+b-1}{2}]$ or $U[\frac{a+b+1}{2}, ..., b]$, as illustrated in Panels A and B of Figure 1.

To further simplify, we assume that D equals a power of 2, and thus after n binary signals the uncertainty is over $D/2^n$ states. The physician chooses the number of mental signals n^M and external signals n^E to obtain a final diagnosis, where

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

Cost of information. The physician faces a fixed marginal cost μ^E for each binary external signal. The marginal cost of thinking depends on available cognitive resources which are determined by the mental load of other activities (in our setting, other patients that the physician is seeing). Let the marginal cost of mental signals be $(L + n^M)$, where L is cognitive load and n^M represents the increasing marginal cost of thinking more about the current patient. The total cost of information acquisition is thus equals:[9]

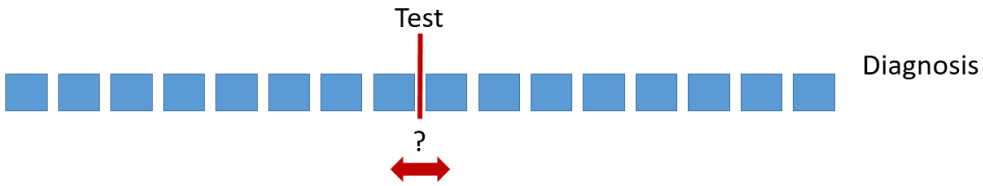
$$C(n^M, n^E) = n^E\mu^E + n^M\Big(2L + n^M + 1\Big)/2. \tag{2}$$

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 $(D/2^{n+1})$.[10]

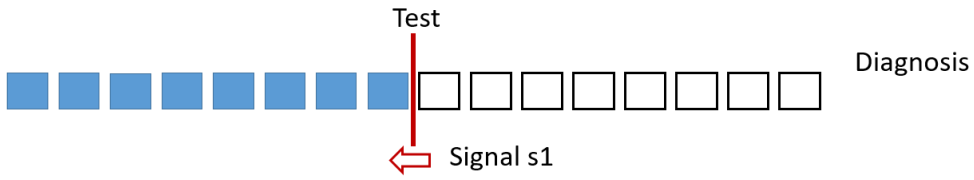
The physician therefore solves the following optimization task to maximize the objective that

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

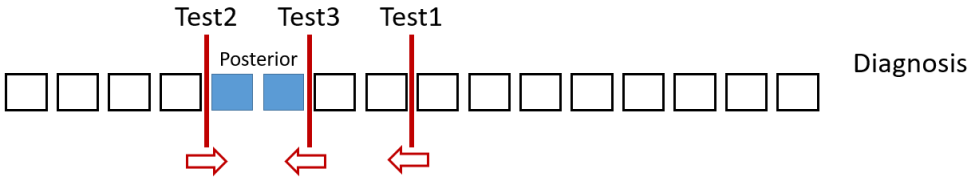
[10]This is because width of the posterior is $D/2^n - 1$. Note that y is set at the posterior expectation. Thus, the expected distance $|y - d|$ of state d from the posterior expectation is expectation over $\{\frac{1}{2}, \frac{3}{2}, ..., \frac{D/2^n-1}{2}\}$.



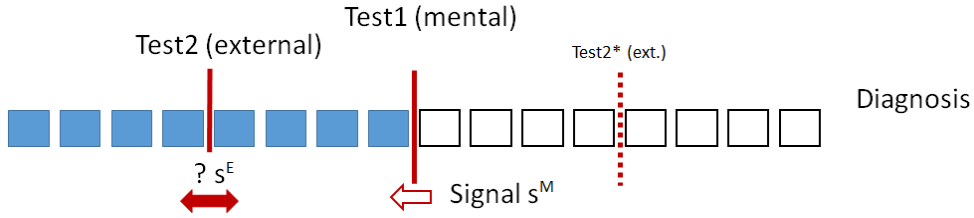
Panel A: Uniform prior belief f^0, $D = 16$.



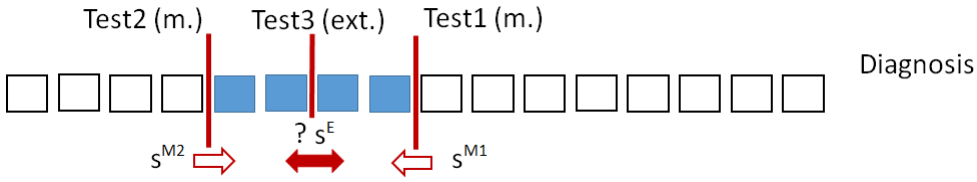
Panel B: Belief f^1 after one signal with a realization of "bottom half."



Panel C: Belief f^3 after three signals.



Panel D: Belief f^1 if the first external signal is Signal 2.



Panel E: Belief f^2 if the first external signal is Signal 3.

Figure 1: Belief distributions across signals/tests.

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. The sequence of the model is thus as follows:

1. Patient of an unknown state d checks in; physician has prior knowledge f^0 about d.
2. Physician chooses how much to think about the patient, n^M.
3. Physician chooses how many diagnostic tests to order, n^E.
4. Physician's knowledge is refined to f^n; physician makes treatment decision $y = E[d|f^n]$.
5. Diagnosis d is revealed.

2.2 Solution and Implications

Under this model set-up, we can derive several results, which we will test empirically. (All proofs are in Appendix A.) We note that all monotonicities referred to in this section are weak ones, which is due to the discrete nature of the choice problem.

Proposition 1. ***(Thinking vs. doing)*** *If prior load L increases, then the physician retrieves a lower number of mental signals n^M (thinks less) and a higher number of external signals n^E (orders more diagnostic tests).*

Proposition 1 has two immediate implications for the degree of unresolved diagnostic uncertainty at different points in time.

Corollary 1. ***(Uncertainty at first diagnostic order)*** *If prior load L increases, then uncertainty at a moment of the first diagnostic order increases.*

The fact that marginal cost μ^E is constant implies the following:

Corollary 2. ***(Uncertainty at admission)*** *Uncertainty in f^n at a moment of the admission decision, after all information is acquired, is independent of load L as long as $n^E > 0$.*

We can further derive implications for the type of test a physician will order, depending on their cognitive capacity. Since cognitive load decreases mental effort at the time of ordering the first diagnostic test (Corollary 1), the first test will be less tailored to a specific d if the physician carries a higher cognitive load. To formalize this prediction, we define the *commonality* of a test as the frequency with which it is ordered. Higher commonality means that the test is used more often and for more types of diagnosis d, i.e., it is more common. For instance, in Panel C of Figure 1, Test1 is more common than Test2, which is more common than Test3. While Test1 is used at the beginning of information acquisition process and thus for 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 are applied at a moment when the physician knows less.

Corollary 3. ***(Commonality of tests)*** *If prior load L increases, then commonality of the first external diagnostic test increases.*

Finally, the model has implications for how to infer physician's beliefs from what diagnostic tests she orders. First, before she orders any tests and before she starts thinking about a patient, her prior belief f^0 coincides with the overall distribution of d. Prior belief can be inferred from the empirical distribution of d as long as the physician has rational expectations.

The same approach applies within our model to any moment when a physician orders an external test. Corollary 3 already introduces connection between the level of thinking and frequency at which particular tests are used. However, the form of beliefs can be identified in more detail: since the physician has rational expectations, then belief at a moment of ordering a test equals the empirical distribution $g(d|test^E)$, where $test^E$ is some particular external test. Panel D of Figure 1 depicts the distribution of diagnoses d under consideration if the first external test chosen is Test2. We can infer that the after mental signal (Test1) pointed to lower half of states. That is, whenever a physician orders Test2 (binary partition between first and second quartile of tests), they reveal ex post that belief f^1 after processing the first mental signal from the patient's chief complaints etc placed the diagnosis d is in the lower half of the distribution only. If on the other hand, the mental signal had pointed to the upper half of the distribution of possible diagnoses, the physician would have ordered Test2*, indicated with the dotted vertical line.

As an additional illustration of the mechanism, consider the situation depicted in Panel E of Figure 1, where the physician chose to think more before ordering a test. Here, her belief after two mental signals – Test1 and Test2 – is uniform over the second quartile, and thus external Test3 is ordered, which partitions the second quartile. In other words, a physician orders Test3 first if and only if mental signals point towards second quartile of states.

More generally, we can partition the distribution of d under consideration from the physician's choice of diagnostic test:

Corollary 4. ***(Inference of beliefs)*** *Belief f at a moment of ordering $test^E$ coincides with a distribution $g(d|test^E)$.*

2.3 Discussion of Assumptions

Form of information. The model assumes a very particular structure of information acquisition based on binary partitions. However, none of the results in Proposition 1 or its corollaries seem to hinge on details of the signal structure.[11] The main driver is that different sources of

[11] There could be minor deviations from Corollary 2 due to discreteness of choices in case prior distribution of d were not uniform.

information, i. e., mental and external, are substitutes and their relative costs affect choices. The results would break only if all mental as opposed to external signals resolved completely different aspects of uncertainty about d and thus the two sources of information were not substitutes at all.

Cost of information. Proposition 1 relies on the assumption that the cost of additional mental signals is increasing, and more so than cost of additional external signals. The same holds for Corollaries 1 and 3. Corollary 2 rests instead on the assumption that marginal cost of external signals is constant. It seems intuitive that while mental efforts face increasing costs, cost of laboratory tests etc is largely independent of prior information acquisition.[12]

Richness of tests. Corollaries 3 and 4 rest on the assumption that there are various types of tests that can be used and are in fact useful in different situations. If, isntead, physicians could only run the same external diagnostic test over and over again, e. g., measurement of blood pressure, to gradually decrease noise, then an outside observer could not infer beliefs from which type of test is run.

To be able to infer beliefs, we need sufficient richness of tests that are available and optimal under different beliefs about d. For instance, the literature on rational inattention assumes full richness, i. e., agents can choose any signals they wish, and thus beliefs can be fully identified from actions.[13] If, instead, the same test were ordered in various situations, then an empirical distribution of d conditional on the test would only identify *average* beliefs across these situations weighed by their relative frequency.

Our empirical setting likely does not provide full richness of tests. However, the richness is still large and probably larger than in most if not any prior empirical study. We observe *how many* different diagnostic tests, and thus this is the number of beliefs across collections of situations between which we can distinguish. Moreover, we pay special attention to beliefs at a moment when the first diagnostic order is placed. This is important because then we identify only beliefs f^{n^M} after thinking before the first order is made, as in Corollary 1, and do not average over potentially very different situations across the whole encounter.

Effect of prior cognitive load. In our model this enters naturally as an increase in information cost, $L + n^M$, similarly as an effect of early mental signals within the same encounter. Alternatively, one could envision that cognitive load could affect physicians prior knowledge, risk aversion or the level of stakes. However, none of the assumptions would generate the substitu-

[12] Moreover, we did not have much intuition for whether marginal cost of binary refinement via external signals of a narrow range should be lower or higher than of a wide range of diagnosis. Such assumptions are in line with our empirical results in Section XXXX

[13] There, a fairly general result is that each action in fact optimally corresponds just to one form of realized posterior knowledge. Actions thus identify knowledge, which can be inferred from the joint distribution of states and actions(Matějka and McKay, 2015; Caplin and Dean, 2015).

tion result as in Proposition 1. Higher prior uncertainty, risk aversion or stakes would (weakly) intensify both mental as well external information acquisition.

Admission decision and non-uniform stakes. Our basic model assumes uniformity of losses across diagnoses. What we neglect is that mistreatment of some diagnoses can be much more costly than of other ones. Moreover, we also do not model a major decision within emergency rooms - and that is whether to admit patients to inpatient care or whether to discharge them. Adding of two interacting elements complicates the analysis a bit, but it also delivers interesting implications.

Consider the following extension to our baseline model: 1) Let treatment $y \in \{1, ..., D, \mathcal{A}\}$, where $\mathcal{A}$ means "admit." 2) The physician's utility from an admission decision equals a constant $\lambda < 0$ irrespective of patients' true d. It is safer to admit a patient to the hospital, but costly. 3) Utility from not admitting but treating a patient in ER equals $-r(d)|y - d|$, where $r(d)$ is a non-decreasing function of d. Higher d's represent diagnosis with higher stakes.

In such a model, the new implications would be that i) patients with higher d's, i. e., higher stakes, are admitted more often; ii) cognitive load increases probability of admission; and iii) uncertainty about d can for admitted patients be both higher or lower than for non-admitted. The intuition for these three results is as follows: The admission decision $\mathcal{A}$ with a fixed loss becomes more appealing if higher d's are more likely, because admission provides hedge against high losses from mistreating higher d's, which implies i). Result ii) emerges because cognitive load increases costs of information, and thus it also increases the motivation to hedge with admission as it decreases the need to acquire costly information. Finally, for iii), there are two competing forces. First, higher d's face higher stakes, which increases information acquisition. On the other hand, according to i) patients with a higher d are more likely to be admitted as a hedge, which decreases the need for information acquisition. Which force dominates depends on $r(d)$ and λ.

Connection to data. The key elements of our model relate to our data as follows: We observe directly the ex-post true diagnosis d as well as treatment decisions y, where the latter include both the admission decision and non-diagnostic orders during an encounter. Prior beliefs are inferred from the overall distribution of diagnoses given an initial chief complaint, while intermediate beliefs are inferred from the modified distribution of diagnoses after also conditioning on which diagnostic tests are used. We describe our measures of current and prior cognitive load, F and G, which capture different elements of L, in Section 4.3.

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 July 1, 2017 and June 30,

2019.[14] Our primary dataset comprises 76,001 visits to the ED, which we refer to as *encounters*.[15][16] Figure 2 illustrates a simplified version of the ED workflow from the patient perspective. Appendix B describes this workflow in detail. We describe different parts of this workflow, when relevant to our analysis, as we proceed through the paper.

Upon arrival (via ambulance or walk-in), the typical patient flow has two main steps prior to being assigned to a physician. First, a registrar collects patient demographic information, solicits information on why the patient has come in, and opens the encounter in Epic. Second, a licensed triage nurse performs a focused assessment in a dedicated triage bay (or at the bedside for direct-to-room arrivals). Vital signs, symptom onset, a pain score and a brief history are obtained and recorded, together with any required isolation flags. Using the Emergency Severity Index (ESI, version 5) algorithm the nurse assigns an acuity code from 1 (immediate life-saving intervention) to 5 (no resources anticipated). These scores are subsequently used for triage and physician assignment, as discussed in the next subsection. The triage nurse also selects a structured chief-complaint label from Epic's menu (e.g., "chest pain", "abdominal pain") or enters free text. There are 732 possible chief complaints at the most granular level of Epic's hierarchy, which roll up into 72 categories of chief complaints in an intermediate tier and 15 categories at the most aggregated tier.

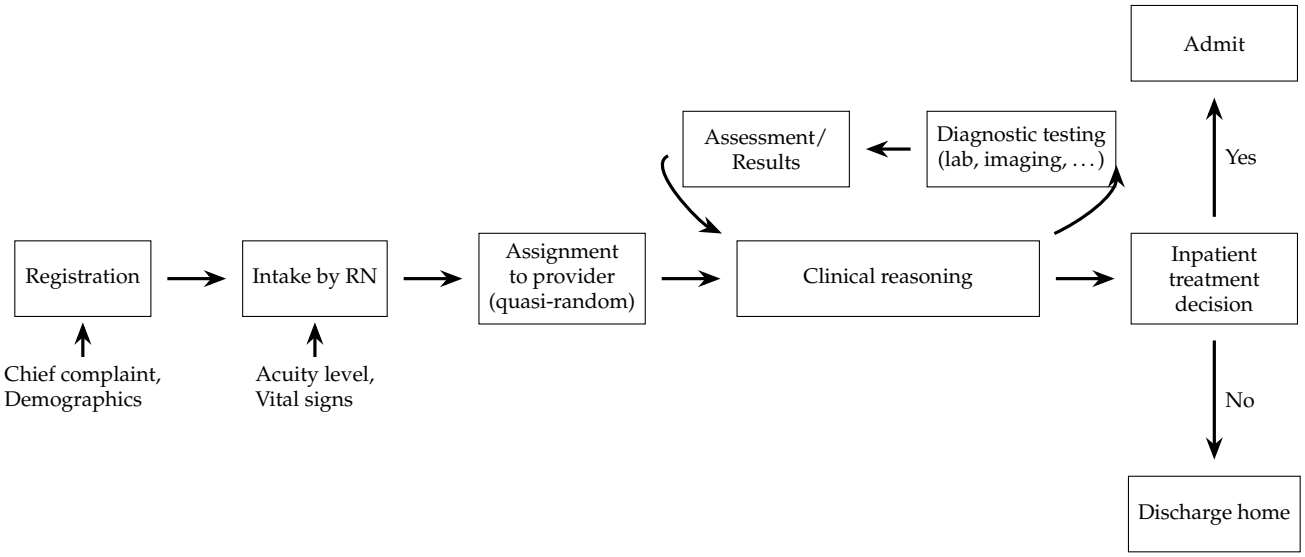


Figure 2: Emergency-department diagnostic and disposition flow.

Table 3 provides summary statistics on patient demographics, health status, and chief com-

[14] Due to privacy restrictions, we don't study minors. Adults comprise the vast majority of ED patients in our context.

[15] 38.19% of encounters involve patients with multiple visits during our observation period, yielding 46,976 unique *patients* in our sample. For simplicity, we refer to encounters and patients interchangeably.

[16] Our data overall have 84,214 encounters but 8,213 are dropped from our primary sample because they have no orders or notes. These patients are either sent home at triage or admitted to inpatient immediately.

plaints. 50.6% of patients are male, 46% identify as white, 15% as black, and 18% as Asian. The mean (median) age is 51.3 years old (52) with a distribution that is left-skewed and covers patients from 18 to the top-coded value of 89. 31% of patients arrive via an emergency vehicle with 69% walking in (or driving in) on their own.

The mean (median) patient spends slightly under 7 (5) hours in the ED. Approximately 76% of patients are discharged from the hospital at the end of their ED stay, while 24% are admitted to inpatient. The five most common patient chief complaints at intake (for the most granular tier) are abdominal pain (10.2%), chest pain (6.5%), shortness of breath (5.7%), a fall (2.9%), or altered mental status (2.8%). There is a small probability of many of the 732 possible chief complaints: we present the top 25 chief complaints, with associated probabilities, in Appendix Table D.4. Appendix Table C.2 describes the rolled up probabilities of chief complaints at the more aggregated tiers (Tier I and Tier II). 61.8% of patients are given an ESI triage score of 3, in the middle of the severity range, with 20.9% given a higher severity level of 2 of 0.4% given the highest severity score of 1. 33.1% of patients have abnormal diastolic blood pressure measured at intake, 5.7% have abnormal respiration, and 3.1% have abnormal pulse.

Table 1: Summary Statistics

	Min	Mean	Median	Max
Panel A: Encounter-level				
Demographics				
Share Male		0.505		
Share Asian		0.178		
Share Black		0.152		
Share White		0.457		
Age	18	50.863	51	89
Health status				
Share admitted to hospital within last 60 days		0.435		
Share ED visit within last 60 days		0.212		
Share upgraded to inpatient		0.238		
Share readmitted within last 30 days		0.039		
Share arrived in emergency vehicle		0.298		
Charlson comorbidity index	0.000	0.674	0.000	16.000
ED length of stay (hours)	0.017	6.981	4.933	125.100
Inpatient length of stay (days)	0.000	5.860	3.875	25.000
Chief-complaint shares				
Abdominal Pain		0.102		

continued on next page

Table 1: Summary Statistics (*continued*)

	Min	Mean	Median	Max
Chest Pain		0.065		
Shortness Of Breath		0.057		
Fall		0.029		
Altered Mental Status		0.028		
Acuity-level shares				
Immediate		0.004		
Emergent		0.209		
Urgent		0.618		
Less Urgent		0.155		
Non-Urgent		0.012		
Missing		0.003		
Abnormal-vital-sign shares				
Abnormal systolic blood pressure		0.107		
Abnormal diastolic blood pressure		0.331		
Abnormal temperature		0.004		
Abnormal pulse oximetry		0.073		
Abnormal Respiration		0.057		
Abnormal pulse		0.031		
Physician actions				
Number of all orders	1	8.136	6	64
Number of diagnostic orders	0	4.552	3	47
Number of medication orders	0	2.321	2	34
Number of non-DX proc. orders	0	1.263	1	40
Number of consults	0	0.252	0	13
Time spent editing notes (in minutes)	0	22.587	13.700	379.667
N = 76,001				
Panel B: Physician-level				
Demographics				
Share Female		0.535		
Types				
Share Resident years 1/2		0.489		
Share Resident years 3+		0.270		

continued on next page

Table 1: Summary Statistics (*continued*)

	Min	Mean	Median	Max
Share Attending physician		0.192		
Share Nurse practitioner		0.049		
Encounter-weighted types				
Share Resident years 1/2		0.419		
Share Resident years 3+		0.262		
Share Attending physician		0.145		
Share Nurse practitioner		0.174		
Shifts				
Number of shifts	1	32.505	11	390
Average shift length (hours)	5.165	11.951	10.824	29.178
Average patients per shift	1.000	2.919	2.278	10.585
N = 610				

Notes: Observations in Panel A are at the *encounter*-level; observations in Panel B are at the *physician*-level.

Information on the 610 medical providers is presented in bottom panel of Table 3. 53% of providers are female. 19% are attending physicians, who are more senior physicians who typically specialize in emergency medicine and who have already completed their residencies. 28.1% of providers are advanced residents, i.e., physicians in years 3 and 4 of their post-medical school training, typically also as emergency medicine specialists. 49.6% of providers are early-stage residents in years 1 and 2 of their residency, while 3.2% are nurse practitioners. The average physician in our sample has 32 shifts, with a heavily right-skewed distribution. Mean (median) shift length is 12 (11) hours and the mean (median) primary patients per shift is 2.9 (2.2), also with a right-skewed distribution. Appendix Table D.1 presents information on the number and distribution of shifts, by physician type, showing similar distributions of shifts lengths by type but each attending and nurse practitioner having more shifts on average.

The Epic Audit Log records every interaction of physicians with the Electronic Medical Record (EMR), allowing us to observe the timing (to the second) and nature of all physician actions, including placing orders (medications, lab tests, etc.), note taking, and inpatient admission decisions. The encounter-level part of Table 3 describes the typical occurrence of different EMR actions for an encounter. For example, mean number of orders per encounter is 8.1, mean diagnostic orders is 4.5, number of consults with other expert physicians is 0.25, and mean time spent editing notes for a patient is 22.6 minutes.

For our primary analysis dataset, we construct a panel at the *patient-provider-action* level, where

each observation corresponds to an interaction between a physician and a patient. We specify a range of actions but focus the most on placing orders, taking notes, and the patient admission / discharge decision. We note that, in many cases, orders are placed in batches, and we account for this in many of our specifications, as described in detail later. For orders, we observe granular details on the type of order placed, the results of the order if diagnostic, and when results are viewed. For notes, we observe metadata on note edit time, note length, and when notes are viewed, but do not observe the full notes themselves for privacy reasons.

3.1 Assignment of Patients to Physicians

Once intake triage process is complete, patients are either taken straight to an available bed ("direct-bed") or returned to the waiting room. Patients are placed into a queue where their position is managed by the charge nurse and displayed to all physicians on a live computerized tracking board (shown in Appendix Figure **??**). The board lists, in order of arrival within each acuity tier, the chief complaint and the elapsed waiting time.

Physicians pick up cases sequentially ("next-up") from this board, typically prioritizing the patient with the most severe triage score who has been waiting the longest. At the moment of assignment physicians observe only the chief complaint and acuity code. Overall, this workflow means that i) initial assignment of acuity level (1 to 5) and chief complaint is nurse-driven and completed before any physician becomes involved and ii) physician-patient pairings are governed almost entirely by the arrival sequence of patients and general physician capacity, rather than by any patient characteristics beyond the triage score. These features are both important for identifying the effects of cognitive load in our environment, as discussed in more depth in the next section.

3.1.1 Quasi-Random Assignment

As an input into our upcoming identification discussion, we assess the quasi-random nature of patient assignment. We do this conditioning on ex ante observable variables (time-of-day, day-of-week, chief complaint, acuity code), which we also condition on in our primary empirical analysis. We define this set of conditioning variables as $\mathbf{T}_i$. We investigate the extent to which observable (ex ante and ex post) patient characteristics, defined as $\mathbf{X}_i$, are balanced across physicians conditional on $\mathbf{T}_i$.

Our main specification uses the propensity of each physician to order diagnostic tests (intensity of treatment) as the physician trait to assess assignment balance with respect to patient characteristics.[17] We define a leave-out measure of the propensity of each physician to order diagnostic

[17]We choose this propensity to order diagnostic tests because (i) it is a one-dimensional measure that distinguishes important aspects of physician differences (ii) we show meaningful cross-sectional differences in this measure across physicians later in the paper and (iii) diagnostic orders placed are a primary outcome variable in our main empirical

tests:

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

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. We then run the regression:

$$Z_i = \alpha + \mathbf{X_i} + \mathbf{T_i} + \varepsilon_i \tag{5}$$

with the primary goal of showing limited variation in Z_i with respect to X_i conditional on T_i. We run this regression using all patient encounters in our sample.

Figure 3 presents the results. First, we run the same regression in panel A, but with number of orders for that specific encounter as the outcome variable, to show that the variables in $\mathbf{X_i}$ capture significant variation in intensity of patient treatment. Next, we show the results from the main balance regression in panel C. Note that, for exposition, we standardize the variables in $\mathbf{X}_i$ to have mean zero and standard deviation equal to one. 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 Z_i. Further, there is no obvious relationship between the coefficients in panel A and panel C. Overall, we see statistically significant but economically small / negligble imbalance in patient characteristics, show by the small F-statistic (5.9) and small R^2 (< 0.005) in (C), while these are much larger in the regression for the number of diagnostic orders in the encounter (F of 406, $R^2 = .18$.).

Panel B in Figure 3 assesses balance with respect to the point-in-time cognitive load measure used in our primary specifications, defined in the next section. We discuss this in more detail next, but note here that balance looks similar for this cognitive load measure and the Panel C leave-out order propensity measure just discussed.

We also note that, while balance tests support the use of conditional random assignment in our main regressions, most of the variation we document in cognitive load comes from the stochastic nature of patient flows into and out of the ED, rather than from assignment once in the ED. We describe this in detail momentarily.

4 Cognitive Load: Measurement and Quasi-Random Variation

A key contribution and key input into our empirical work is identifying and measuring granular point-in-time cognitive load measures for physicians. Additionally, a key goal is to establish the

analysis.

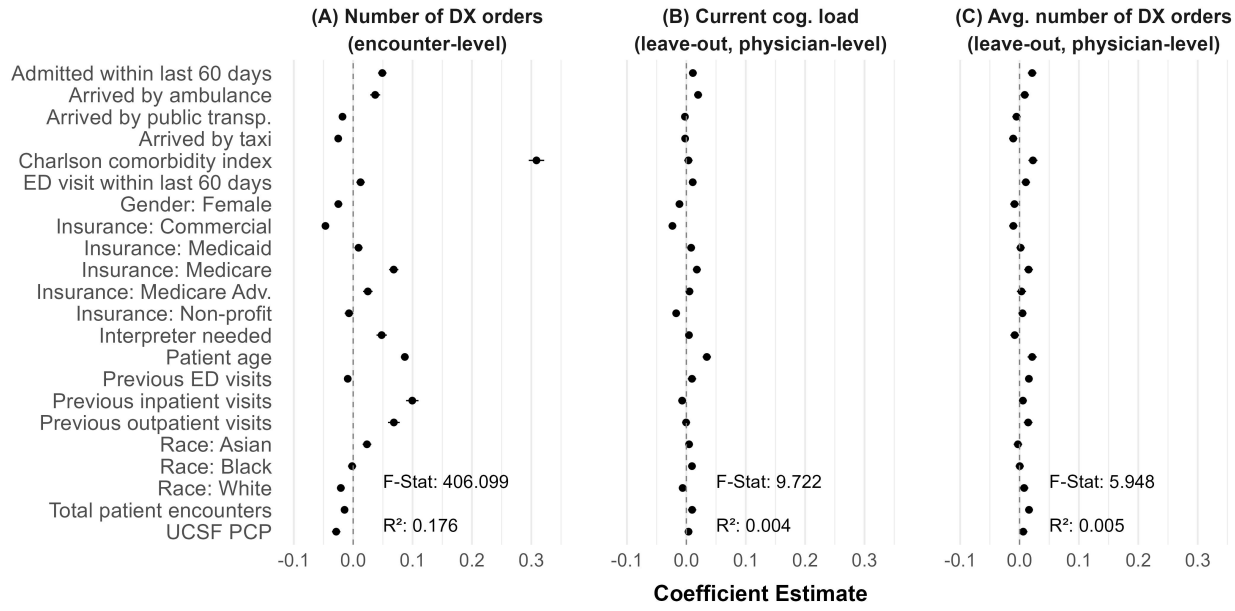


Figure 3: Covariate Balance

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. 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.

existence of meaningful quasi-random variation in cognitive load to support our main empirical analyses. We present our approach and address these points in this section.

4.1 Identification

Empirically, we resolve two key identification issues in order to establish the impact of cognitive load on different outcomes. These two challenges are:

1. **Endogenous Observed Outcomes and Measuring Cognitive Load.** The interim and final outcomes we observe occur for physicians with different levels of cognitive load and are therefore endogenous to that cognitive load. At the same time, we want to base our measures of cognitive load on the number of and intensity of patients a physician is treating / has treated. Consequently, we can't use actual post-physician-assignment actions as inputs into our measures of cognitive load.

   We address this issue by forming measures of predicted patient intensity using only information that is ex ante to provider assignment and treatment. We use predicted patient intensity as an instrument for actual intensity when forming our physician-time-specific measures of cognitive load, so that these measures, which relate to the typical mental effort used for a given patient, are not based on endogenous inputs. We describe this prediction process in detail in the next sub-section.

2. **Quasi-Random Arrival, Assignment and Variation in Cognitive Load.** The second identification challenge we address is ensuring that we have quasi-random variation in cognitive load, within and across physicians. We leverage both quasi-random patient arrival to the ED as well as quasi-random assignment of patients to physicians once they have arrived in the ED. Later in this section, we present a range of evidence showing and describing significant within-physician quasi-random variation in cognitive load over time.

4.2 Measuring Expected Workload Intensity for each Encounter

For each encounter, we construct a measure of the expected workload intensity c_i based on three outcomes: (i) the number of orders (ii) the log of note edit time (iii) the number of consultations obtained with other expert physicians. As described in the next subsection, we use these predictions of expected workload intensity as inputs into our point-in-time measures of cognitive load for each physician.

We predict each of these three variables using cross-validated LASSO regression, leveraging encounter-level variables that are ex ante to physician assignment and treatment. These predictors include patient age, patient vital signs at triage intake, the three different tiers of chief complaints (including the lowest tier with 732 possibilities), the ESI triage score, means of arrival to the ED,

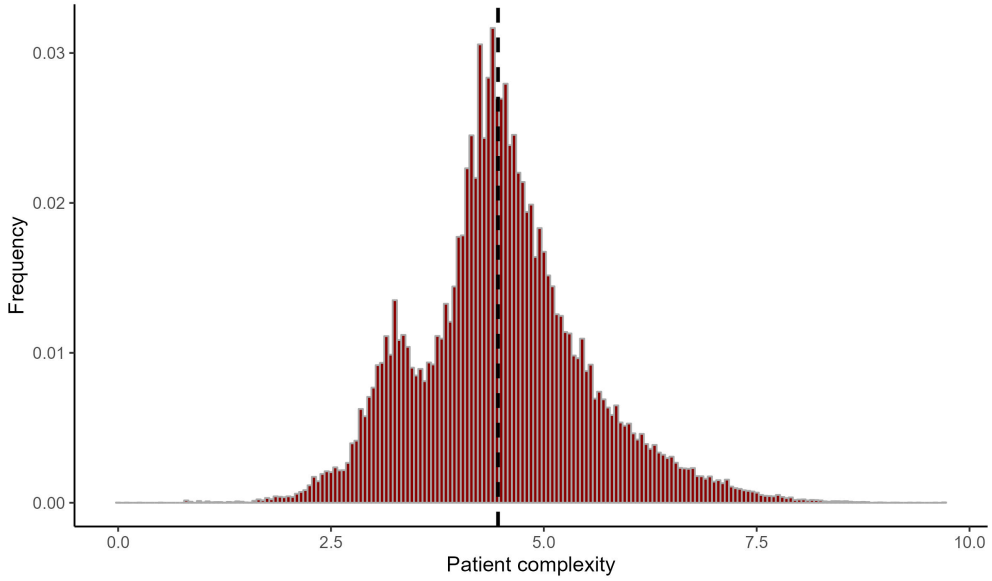


Figure 4: Distribution of Patient Complexity Measure

Notes: The figure shows the distribution of the patient complexity measure for the full-sample specification using all three chief complaint categorical variables as predictors: Tier 1 (T1), Tier 2 (T2), and detailed (CC) chief complaint. The vertical dashed line shows the median.

recent ED / admission history at UCSF, the need for an interpreter, and the Charlson co-morbidity index. Methods for dealing with sparse predictor matrices are important for our process given the importance of chief complaints in prediction. Appendix C.1 describes our LASSO procedure in full detail including myriad sensitivity checks, measures of goodness of fits, and split-sample tests to ensure we are not overfitting our predictors. Additionally, we analyze alternative prediction specifications, i.e. gradient-boosted trees and neural networks, finding that LASSO delivers similar results in our context.

Appendix Tables C.5-C.7 show the LASSO coefficients of the 25 most important predictors for each of the three outcome variables. ESI triage acuity codes and distinct chief complaints are important predictors for the all three outcome variables, with chief complaints impacting each outcome variable in sensible ways based on the complaint and outcome variable. Though, as expected, there are some predictor differences across these three distinct outcomes, predictions are highly correlated across them.

For each encounter we combine the three predictions into a composite measure c_i by summing the standardized predicted values. This composite score is further normalized to a z-score. Finally, for ease of interpretation, 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 4 shows the distribution of patient complexity measures, which has two peaks and is right-skewed.

4.3 Cognitive Load Definition

We use these patient intensity predictions as inputs into our measures of physician point-in-time cognitive load. We compute point-in-time cognitive load for physician j taking an action for patient i at time t as follows: We define cognitive load as the sum of the encounter-level complexity measures $c_{i',j,t}$ of all distinct patients excluding the focal patient i ($i' \neq i$) for whom the physician has taken an action (i.e., orders, notes, admission decision) during the preceding 90 minutes within the same shift. We show in Appendix F.3 that all main results are robust to this assumed time window, and hold, e.g., when using instead a 60-minute or 120-minute window.[18] Note that, 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.

Physician j's cognitive load $F_{i,j,t}$ when attending to patient i at time t is thus given by:

$$F_{i,j,t} = \sum_{i' \in \mathcal{I}_{j,t} \setminus \{i\}} c_{i',j,t} \tag{6}$$

where $\mathcal{I}_{j,t}$ is the set of patients with whom the physician j has interacted at least once within the last $t - 90$ minutes in the current shift. Note that, if physician j has been on shift for less than 90 minutes as of time t, $\mathcal{I}_{j,t}$ comprises all patients with a physician action since the beginning of the shift.

Figure 5 presents the distribution of our measure of point-in-time cognitive load across all patient order actions in our main sample (266,625 observations). This shows significant variation in cognitive load, and also highlights the right-skewed nature of this distribution.

4.4 Variation in Cognitive Load

Figure 6 illustrates the variation of point-in-time cognitive load along several dimensions. Each figure reports different measures related to cognitive load (y-axis) as a function of hour-in-shift for a physician (x-axis).

Panel A in Figure 6 presents distribution of within-physician averages for cognitive load as a function of shift-hour. Thus, for a given physician, the figure averages point-in-time cognitive load for a given shift hour and reports the values for each physician (with a physician-specific line) across all shift hours. For expositional purposes, the figure shows this for the top 20 (by number of patients) for advanced stage residents. As expected, the figure shows a tight distribution of these average values at each hour across physicians, highlighting the quasi-random nature of patient assignments within shift and over time.

We decompose variation in cognitive load within and across physician to highlight the stochastic nature of this variation. For example, during the fifth hour the shift, physicians have an av-

[18] Additionally, we show that this definition is robust to including patients who the physician is assigned to, who are still in the ED, but for whom no action is taken in the prior 90 minutes.

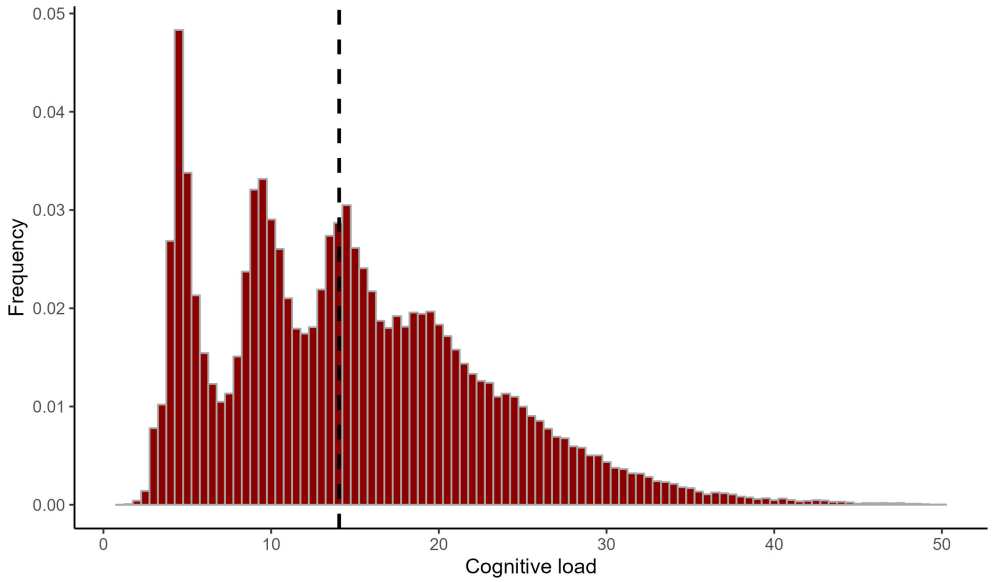


Figure 5: Distribution of Cognitive Load

The figure shows the relative frequencies of the cognitive load measure with a binwidth of 0.5. Each observation is the cognitive load at one *encounter-order action*. The vertical dashed line indicates the median.

erage concurrent cognitive load of 8.60, with a total standard deviation of 4.76. Decomposing this overall variation via the Law of Total Variance[19], dispersion in this hour is driven largely by within-physician fluctuations – 76.2% of the total variation – while systematic differences across physicians account for only 23.8% of the variation. This relationship holds across shift hours. Additionally, much of the systematic variation present is due to effects like day-of-week or time-of-day, which we condition out in our primary analyses. This strongly underscores the main identifying assumptions in our regressions, namely that cognitive load is random conditional on the many control variables included.

Panel B highlights this further with an illustrative example of cognitive load for 10 shifts of one advanced resident. This panel highlights typical across-shift variation in cognitive load for a given physician, which is notably larger than any systematic differences noted in the statistics above and in Panel A.

Panel C provides an even more granular example, highlighting the leave-out cognitive load trajectories for two ex ante similar patients being treated by one resident on different shifts. The two patients come in at the same time of day and day of week, have the same chief complaint and acuity triage scores. Thus, the conditioning variables for our primary regressions are the same for these two patients, with the only differences being changes to the cognitive load of the same treating physician. The figure highlights quite different leave-out cognitive load for the physician

[19] $Var(F) = Var(E[F \mid J]) + E[Var(F \mid J)]$, where the first term represents the *between*-component, and the second term the *within*-component.

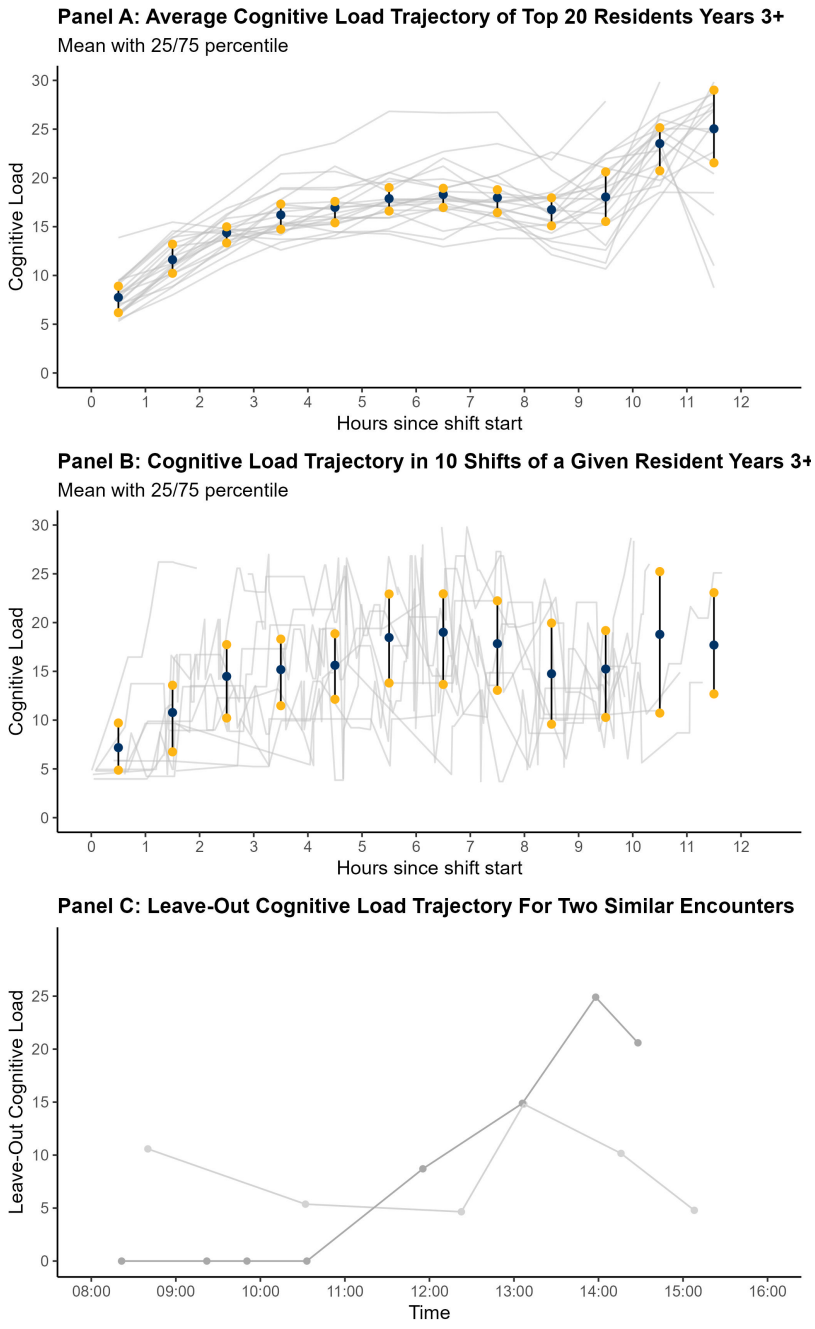


Figure 6: Trajectory of Cognitive Load

Notes: Panel A shows the average cognitive load at each shift hour for the 20 providers with the most shifts who are in year 3 or later of their residency. The blue dot denotes the average while the bottom yellow dot denotes the 25-th percentile and top yellow dot denotes the 75-th percentile of cognitive load across all shifts.
Panel B shows the cognitive load for 10 randomly selected shifts of a *given* resident in year 3 or later. These shifts are selected from the 37 shifts for that provider that last between 10 and 12.5 hours.
Panel C traces the *leave-out* cognitive–load faced by the treating physician during two abdominal-pain encounters that began on a Friday between 08:00 and 09:00 and were both triaged as *Urgent*. In each case the provider was a resident in the second hour of a 10-hour shift, yet the intensity of concurrent workload diverges sharply. The solid line shows the cognitive load for the focal physician after removing the focal patient from the calculation. Dots indicate order "batches."

over the course of each patient encounter, highlighting the exact variation we use in our upcoming analyses.

We also assess how variation in cognitive load comes from two proximate determinants – patient head-count versus case-mix complexity. We find that 86.3% of the spread is explained by the number of simultaneous patients, whereas 13.7% stems from how clinically demanding those patients are. This highlights that the stochastic nature of the number of patients flowing into or out of the ED is a key factor underlying variation in cognitive load in our setting, while variation in case complexity conditional on number of patients provides some additional variation.

4.5 Prior Cognitive Load

We also measure prior cognitive load from earlier in a shift in order to aggregate cognitive-load effects. Here, we sum up 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:

$$G_{i,j,t} = \sum_{i' \in \mathcal{I}_{j,t-90} \setminus \{i\}} c_{i',j,t} \tag{7}$$

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 with patient i.

We note that there are a variety of ways to account for prior cognitive load, including (i) measures that integrate current cognitive load with some depreciated notion of past cognitive load, or (ii) time-series lags for prior 90-minute windows. Additionally, we note that we could also incorporate longer-duration measures related to cognitive load experienced within other shifts over the past week or month. Finally, in a coarse way, one can think of the prior cumulative cognitive load measure as a reflection of behavior dynamics that occur when cognitive load shifts in specific directions over time.

4.6 Discussion

We define cognitive load based on the number and intensity of patients a physician sees at a given point in time. As discussed in the cumulative measure above, these point-in-time measures can be combined over time in different ways to reflect a broader notion of cognitive load that depends on the past and present patient burden.

While the number of patients and number of actions per patient are a natural starting point for mental burden, there are a number of other possibilities that could extend this notion. First, it is plausible that even if a patient has many orders and note actions, that something about the treatment experience is very routine and, consequently, does not pose a larger mental burden then a patient with a less routine case but few expected orders or note actions would. Second,

one could imagine that seeing many complex patients of exactly the same kind in one day would pose a lower mental burden than seeing many patients of similar complexity but different chief complaints / conditions. Third, it may be that certain types of patients, e.g. those with dementia co-morbidities, are especially taxing to treat regardless of the primary chief complaint.

We view these kind of extensions as additive to our approach, in the sense that they would make our approach more precise and allow testing of more subtle cognitive load hypotheses. Overall, we find it unlikely that these additions would subvert our primary assumption that treating more demanding patients in terms of expected actions, and a larger number of those patients, leads to higher cognitive load. These kind of extensions could leverage hypothesis such as those mentioned above and those discussed in prior work (e.g., Enke and Graeber (2023) and Oprea (2024)) to study whether specific forms of potential cognitive load have meaningful impacts.

5 Empirical Approach

Our primary empirical analyses come from a series of regression models with the following form:

$$Y_{i,j,t} = \alpha + \beta f_1\left(F_{i,j,t}\right) + \gamma f_2\left(G_{i,j,t}\right) + \delta_t + \lambda_i + \theta_j + \varepsilon_{i,j,t} \tag{8}$$

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 (6), and $G_{i,j,t}$ represents the leave-out measure of long-run cognitive load defined in Equation (7). We consider specifications where $f_1(\cdot)$ is either a z-score transformation or a set of quintile indicators for the underlying variable. We primarily use z-score transformation for $f_2(\cdot)$ but do at times specify it as a set of quintile indicators.

Our approach focuses on leveraging the within-physician variation in cognitive load described in depth in the last section. To this end, we include physician fixed effects θ_j. We include encounter-level fixed effects for chief complaint and acuity score – captured by λ_i – for the twin purposes of (i) conditioning on the ex ante patient information known at the time of assignment and (ii) to isolate deviations in key outcomes due to cognitive load, above and beyond typical treatment for a given chief complaint and acuity level. We include several different time-related fixed effects including (i) time of day, (ii) day of week, (iii) time into physician shift, and (iv) time until the physician shift ends, captured by δ_t.

In most of our analyses we focus on β and interpret this coefficient as the causal effect of changes in recent cognitive load on our key outcomes of interest. The set of fixed effects allows us to isolate the effect of cognitive load conditional on other contextual factors that matter (time of day, day of week, typical physician practice style, type of patient condition / complaint). We can also use this specification to study how the impact of cognitive load compares to variation generated by the observable factors included in the fixed effects.

In our primary specifications, for simplicity we include only orders and notes from the assigned primary provider. In practice, 97% of notes and 60% of orders are input by the primary assigned provider. We include all orders and notes for all providers for a patient (as if they were assigned to the primary physician) in the Appendix (F.5), finding no change in the direction/implications of our results.

6 Results

One of our primary focuses is to assess how cognitive load impacts physician diagnostic processes. To this end, many of our outcomes focus on different aspects of diagnostic orders, i.e. orders placed with a specific goal of moving towards a final diagnosis. These outcomes relate to comparative statics presented in our model in Section 2.

We begin by looking at the amount of diagnostic orders placed. One comparative static from our model is that the number of diagnostic orders placed will increase as cognitive load increases, as a substitute for physician thinking effort. We use an order batch for patient i and physician j as the action (occurring at time t), and the outcome variable in Equation (8) is the number of diagnostic orders placed in the order batch.

Table 2 presents the results in columns (1) (without physician fixed effects) and (2) (with physician fixed effects). For our preferred specification with physician fixed effects, we find that a one standard deviation increase in cognitive load leads to 0.05 additional diagnostic orders in a batch, equivalent to a 4 percent increase relative typical order amounts. These estimates have a high degree of precision, indicative of the large number of observations in our data.

Columns (3) and (4) report the effect of cognitive load on medication orders, providing a useful validation check since medication orders should be less subject to cognitive load and not closely related to diagnostic effort. We find that a one standard deviation increase in cognitive load has a precise null effect on the number of medication orders. This highlights how physicians are not uniformly increasing all order types and, instead, selectively expanding the number of diagnostic orders.

Columns (5) and (6) show specification results where the outcome variable is the share of diagnostic orders with an abnormal result. This also has a precise null effect, highlighting that, while the number of orders is expanding, there is no change in the proportion of orders delivering abnormal flags. [20]

We note that this specification, and most that follow, occur at the diagnostic order action level. This means that it assesses the intensive margin of orders placed in an order batch. We focus on the intensive margin because it allows us to precisely control for cognitive load at the time of

[20] Note here that since the mean share of orders with abnormal flags is 0.06, and the one standard deviation increase in cognitive load increases the order share by 4%, even if all of the incremental orders had a 0% chance of detecting an abnormality, this estimate could still be a precise zero change in the overall share of orders that are abnormal.

Table 2: OLS Estimates of the Impact of Cognitive Load on Orders

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.071***	0.050***	0.010***	0.002	0.0003	0.001
	(0.006)	(0.006)	(0.003)	(0.003)	(0.001)	(0.001)
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
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.30	1.30	0.66	0.66	0.06	0.06
N	266,625	266,625	266,625	266,625	266,625	266,625
R^2	0.043	0.052	0.024	0.057	0.012	0.017
Adjusted R^2	0.040	0.047	0.021	0.052	0.009	0.012

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result 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%).

order placement rather than relying on an average (or other moments) over the course of a given encounter. As described later in our robustness section, we also assess the effect of cognitive load on the extensive margin of diagnostic orders, i.e. how many distinct order batches a physician places. We find that a physician with a one standard deviation higher leave-out cognitive load, averaged over a patients full encounter, leads to 0.117 more order batches per encounter, a 3.8% increase over baseline.

Additionally, our main specification above assumes a linear effect of cognitive load on the outcome variables. Appendix Table **??** examines a specification where $f_1\left(F_{i,j,t}\right)$ maps our cognitive load measure $F_{i,j,t}$ to a set of quintiles. We leave out the middle quintile and use β to assess the relative effect of being in another cognitive load quintile. We find an ordered, but non-linear effect, finding that most our linear model effect is driven by physicians in the top 20% of cognitive load. That group increases the number of diagnostic orders by 8% relative to when they have median

cognitive load. The top two quintiles have positive effects, while the bottom two quintiles have negative effects, relative to the leave-out middle quintile. As with our average, there is no effect of these higher cognitive load quintiles on medication orders.

A next natural question is to ask how the types of diagnostic orders placed change as cognitive load increases. Table **??** shows a clear shift in the profile of orders. When physicians have higher cognitive load, they place a meaningfully higher number of diagnostic orders in the set of top 25 most common diagnostic orders. In the specification with physician fixed effects (column (2)) physicians place 8% more common orders per order batch for a one standard deviation increase in cognitive load. Column (4) shows that the flipside of this is also true. A one standard deviation increase in cognitive load *decreases* the number of diagnostic orders by 5%. Later, we replicate this finding for specific common chief complaints. These findings highlight how, when physicians are cognitively loaded, they shift to a more "standardized" or "routine" set of diagnostics and move away from more targetd, less common orders, commensurate with a model where physician do less thinking / evaluation of a given patient up front.

Table F.2 studies how the profile of orders changes uses cognitive load quintiles (instead of linear in standard deviation). The pattern is striking: for the most cognitively loaded quintile, common (top 25) diagnostic orders increase by 13% while uncommon orders decrease by 5%, relative to the median leave-out cognitive load quintile. For the least cognitively loaded quintile, common diagnostic orders are 5% less likely and uncommon are 10% more likely. The relationships are monotonic across the five quintiles as cognitive load increases, both for common diagnostic orders and uncommon ones. Additionally, in appendix section F.6, we study the threshold for common / uncommon orders, using the 10 most common, 50 most common, and 100 most common to separate these catergories. We find similar patterns in these analyses to those described here.

6.1 Diagnostic Precision

We now investigate the precision of diagnostic orders, in a structural sense, as a function of cognitive load. We construct a measure of the "informativeness" 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. This links directly to the comparative static in our model, which suggests that, when physicians are cognitively loaded, they exert less mental effort on a given patient up front and start placing diagnostic orders with more diffuse priors.

To calculate our informativeness measure, we proceed in two steps. In the first step, we compute the empirical probability mass function (pmf) of final diagnoses for an individual patient receiving a specific diagnostic orders o_z. Since orders are used differently depending on the chief complaint, we estimate the pmf of final diagnoses for a given order conditional on each chief complaint *cc*:

Table 3: 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.)	0.090***	0.068***	−0.019***	−0.018***	−0.174***	−0.071***
	(0.005)	(0.005)	(0.002)	(0.002)	(0.005)	(0.004)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.046	0.054	0.022	0.032	0.023	0.112
Adjusted R^2	0.043	0.049	0.019	0.027	0.019	0.107

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)). For columns (5)-(6), the outcome variable is log-transformed, while its mean is reported in raw seconds. 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%).

$$\widehat{\Delta}_{DX}\left(o_z \mid cc\right) \tag{9}$$

Next, let $O_i = \{o_{i,1}, \ldots, o_{i,Z}\}$ denote the set of diagnostic orders placed for encounter i. To rule out dynamic considerations such as 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 for an order set O_i, 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:

$$H_i := H\left(\sum_{p=1}^{P} \widehat{\Delta}_{DX}\left(o_{i,z} \mid cc_i\right)\right) = -\sum_{d=1}^{D} p(d \mid O_i,\, cc_i) \log_2\left[p(d \mid O_i,\, cc_i)\right] \tag{10}$$

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 $\sum_{p=1}^{P} \widehat{\Delta}_{DX}\left(o_{i,p} \mid cc_i\right)$. Under this measure, a higher entropy value H_i implies an order set that indicates more a dispersed prior over possible final

diagnoses.

Next, we go back to the ex ante phase and compute the entropy over all final diagnoses associated with each chief complaint cc_i. This provides a baseline entropy for each chief complaint that gets reduced as thinking occurs and orders are placed. Define this ex ante chief complaint entropy as $H(\widehat{\Delta}(cc_i)$. We define the entropy reduction from physician assessment and contemplation, prior to placing the first order batch, as:

$$E_i := H\left(\widehat{\Delta}(cc_i)\right) - H_i \tag{11}$$

Here, E_i measures the ex ante entropy for a patient given the chief complaint and other factors, e.g. acuity code, minus the updated entropy H_i which accounts for physician assessment and contemplation prior to placing the first order batch (while still accounting for the ex ante factors). We use these measures as outcomes in our main specification to assess the impact of cognitive load on mental effort and diagnostic precision.

Table 4 shows the results for the impact of cognitive load on entropy reduction E_i. This analysis focuses on the 56,398 encounters for which any diagnostic orders are placed within 30 minutes of the first physician-patient interaction.[21] The results in column (2) implies that a one standard deviation increase in cognitive load leads to a 2% decrease in the entropy reduction, holding all else equal. Put differently, physicians with higher cognitive load are less precise in placing relevant diagnostic orders, i.e., placing less effort in contemplation pre-order-placing. A 2% decrease in entropy reduction at the mean given in Table 4 (0.62 bits) corresponds to about 0.01 fewer bits of information, making the posterior distribution roughly $2^{0.01} \approx 1.007$ times more diffuse, which is about a 0.7% increase in the effective number of plausible diagnoses.

Column (3), our preferred specification, adds a control for the number of diagnostic orders placed, to account for the mechanical effect that a larger order set is informative over more diagnoses. Since our earlier results show that cognitive load increases the number of orders placed, we believe this is a valuable control. The coefficient of interest shows a slightly larger effect than that in column (2), suggesting that the reduction in diagnostic precision among cognitively loaded physicians is not driven by these physicians simply placing more diagnostic orders.[22] Overall, these results highlight how cognitively loaded physicians are less precise in their diagnostic orders, holding many factors constant (including the physician and chief complaint).

[21] In the appendix, we also assess the impacts of cognitive load on entropy reduction with different initial time windows, finding similar results.

[22] Column (4) adds the share of diagnostic orders in the 100 most common diagnostic orders to account for the potential mechanical effect that more frequently placed orders are associated with a wider array of diagnoses. The coefficient of interest decreases minimally, and we don't prefer this specification because controlling for more or less common orders in any way has a clear relation to our measure of diagnostic precision.

Table 4: 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^{***}	-0.014^{***}	-0.011^{**}	-0.009^{**}
	(0.004)	(0.005)	(0.005)	(0.004)
Number of DX orders			-0.049^{***}	-0.037^{***}
			(0.001)	(0.001)
Share of top100 DX orders				-1.767^{***}
				(0.031)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.62	0.62	0.62	0.62
N	56,398	56,398	56,398	56,398
R^2	0.492	0.506	0.534	0.640
Adjusted R^2	0.485	0.494	0.523	0.631

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 is the encounter, considering 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%).

6.2 Documentation Effort

Before we move on to our analysis of entropy and diagnostic precision, we present one more view into how medical practice changes with cognitive load. We examine documentation effort by examining the time a provider spends editing patient notes. This measure, developed from audit log data, provides time-stamped information on physician note-taking, allowing us to assess whether physicians reduce note-taking when cognitively loaded.

Columns (5) and (6) from Table 3 reports the results from specification (8), using 181,154 note edits as the unit of observation and the natural logarithm of note editing time as the outcome. Column (6) shows that a one standard deviation increase in cognitive load reduces editing time by about 7.1 percent. This suggests that, physicians have less mental bandwidth to edit notes,

though we don't assess whether or not this has positive or negative welfare effects.[23]

6.3 Hospital Admissions

One key outcome from a visit to the emergency department is whether the patient is admitted to the hospital as an inpatient. This is a very consequential and costly decision, often leading to a meaningfully longer and more intensive patient visit. If physicians are cognitively loaded, it is plausible that this could impact this important decision as discussed earlier in section 2.

We investigate this empirical by modifying our main regression specification such that an observation is at the encounter level rather than at the patient-event level. Since the admission decision is a product of many things occurring during the encounter, it is appropriate to treat this outcome as reflecting the many things that happen during the ED encounter. The primary change moving to the encounter-level specification is that instead of using a point-in-time cognitive load measure, we assess cognitive load over the entire patient encounter and use a maximum of cognitive load during that time as our variable of interest.

Table G.13 presents the results. Strikingly, we find that a one standard deviation increase in maximum cognitive load for the physician during the patient encounter increases the change of admission by 9%, with a very high level of statistical significance. As with earlier estimates, this controls granularly for physician, chief complaint, acuity code, and time fixed effects. Columns (3) and (4) in Table G.14 in our next section breaks the max cognitive load variable into deciles at the encounter level and regresses indicators of those details on the binary admission outcome, controlling for the same fixed effects. The effect increases clearly moving from the low to high cognitive load deciles: all else equal, the admission probability of a patient is 28% higher for a physician in the highest cognitive load decile relative to one in the lowest decile.

We also assess the impact of cognitive load on readmission to the ED within 30 days (columns (3) and (4)). We find a precise null effect on readmissions.

The admissions result is notable and bears further discussion. One possible reason for the admission increase is that the ED bed capacity is met when many patients are in the ED and and there is now an additional shadow value of admissions coming from that capacity constraint. The UCSF ED operates with flexible capacity rather than a fixed bed count: during our study period, it comprises 35 patient rooms plus about 25 hallway bays, with observation and triage areas opened or closed as needed. We assess the number of patients in (and moving through) the ED during periods when physicians have high cognitive load. To do this, we plot the total number of patients in the ED when physicians have an order action for a patient and are categorized as being in the top two quintiles of cognitive load. Appendix Figure D.3 plots this distribution. The

[23] As for resource utilization, our analysis focuses on the *intensive margin* of note edit time. 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 and highlighting a more piecemeal production of notes when cognitively loaded.

Table 5: OLS Estimates of the Impact of Cognitive Load on Inpatient Admission and Readmission at the Encounter-Level

	Inpatient admission		Readmission within 30 days	
	(1)	(2)	(3)	(4)
Max cognitive load (std.)	0.023***	0.021***	0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.24	0.24	0.04	0.04
N	76,001	76,001	76,001	76,001
R^2	0.221	0.242	0.048	0.062
Adjusted R^2	0.213	0.228	0.039	0.045

Notes: The table presents OLS estimates of linear regressions of dummy variables whether the patient was admitted to the hospital during the encounter or re-admitted within 30 days on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

figure highlights multiple reassuring findings. First, the total number of patients in the ED at one time (including those not assigned to a bed yet) is typically well below capacity with only 9.66% of observations occurring when the ED is over 80% capacity. Second, and perhaps more importantly, the distribution of patients in the ED for physician actions in the top two quintiles of cognitive load looks very similar to the distribution for the lower three quintiles, suggesting that these kind of structural capacity issues do not cause these results.

A potential model-based explanation for our admissions finding is physician risk-aversion that manifests in cases with high cognitive load. Our precision results above highlight how physicians with higher cognitive load do less contemplation up front and place less targeted orders for patients, all else equal. In Table G.3 in the appendix, we highlight how this result also holds at the encounter level: i.e., when a physician has higher max cognitive load over a patient encounter, entropy reduction over the whole encounter is lower.[24] Thus, one potential underlying mechanism

[24]Note that as in our primary results, though entropy reduction is lower when max cognitive load is higher, the

is that physicians have more coarse information about a patient when they are cognitively loaded and, as a result, they are more likely to admit the patient because they are risk averse about the (higher) probability of a negative outcome if they discharge the patient instead.

6.4 Additional Results

As we presented our results, we highlighted a number of additional specifications and robustness checks in the appendix. These include (i) alternative functional forms for cognitive load in our main regression specification (e.g., indicators for quintiles of cognitive load) (ii) different functional forms for outcome variables (e.g., logs vs. levels) (iii) alternatives to our 90-minute window for defining action-level cognitive load (60 minutes and 120 minutes) (iv) regressions including all provider actions taken for a patient, rather than those of just the primary provider (v) specifications that change the threshold for defining common vs. uncommon orders and (vi) a range of specifications that are based at the encounter level, rather than the encounter-action level. We highlighted these specifications when relevant, in comparison to our primary results earlier, finding that they generally support our primary conclusions and expanding our understanding of the effects in the ways described.

We also study how our analysis applies to specific common chief complaint, to get a more granular assessment of our main findings. For this analysis, we study two common chief complaints where diagnoses is crucial, abdominal pain and chest pain. For abdominal pain, a chief complaint for 9% of the encounters in our main sample, we find very similar results to our primary findings. A one standard deviation increase in cognitive load increases diagnostic orders by 4%, does not increase the number of medication orders, meaningfully increases the number of common orders (9%), meaningfully decreases the number of rare orders (12%), decreases note edit time (9%), and reduces entropy reduction in the first order batch by 3% (though this latter effect is marginally not statistically significant).

For chest pain, a chief complaint for 4% of the encounters in our main sample, we also find very similar results to our primary findings. A one standard deviation increase in cognitive load increases the number of diagnostic orders by 8%, slightly *decreases* the number of medication orders (3%), meaningfully increases the number of common orders (10%), decreases the number of rare orders (4%, marginally not statistically significant), decreases note edit time (8%), and reduces entropy reduction in the first order batch by 3% (though this latter effect is marginally not statistically significant), and does not have enough statistical power to detect an entropy effect below a 10% change (which it does not).

These results for individual chief complaints are consistent with our findings aggregating over all chief complaints. They open a window into future work that leverages structural differences

absolute number of orders is also higher in this case (Table G.1). This continues to suggest that, under higher cognitive load, physicians place more, less targeted, orders.

in the medical practice of treating these distinct chief complaints to (i) model diagnostic pathways in a more sophisticated way (ii) assess the impacts of cognitive load on those pathways and (iii) assess heterogeneity in physician practice styles in a granular way and (iv) assess the interactions between those practice styles and cognitive load. To answer some of these questions effectively, researchers will need an even larger sample of patients than that which we have here, in order to gain statistical power and link structural insights across different categories of chief complaints.

7 Patient Reallocation Analysis

One possible implication of our findings is that, if patients could be allocated to providers with lower cognitive load, we could reduce hospital admissions and reduce the number of diagnostic orders, while actually increasing the informativeness of those orders. In this section, we analyze this possibility, while being cognizant of the fact that we hold many contextual and dynamic factors fixed in order to do so. In that sense, we view this counterfactual analysis as suggestive of the potential for patient reassignment in a world with cognitive load, rather than a full treatment of this problem.

To do this, we leverage estimates from an alternative regression specification that is at the encounter level. As discussed earlier, in this specification cognitive load for an encounter is assumed equivalent to the maximum cognitive load that occurs at a point-in-time during the encounter. As noted earlier, relative to our main event-action point-in-time regressions, this allows for an extension of our framework to the encounter level, while somewhat reducing the closeness in time of the cognitive load measure to the regression outcome of interest. We classify this cognitive load into deciles at the physician-encounter level, and use indicators for those deciles in our specification. We continue to control for the same granular fixed effects.

The results from these specifications are presented in Figure 7 below, with the detailed coefficient estimates presented in the Appendix in Table G.14. As discussed earlier in our section on hospital admissions, the likelihood of inpatient admission increases significantly as cognitive load increases over the deciles (columns (3) and (4)). The relationship is near monotonic, and shows that, all else equal, when a physician is in the highest decile of cognitive load they are 28% more likely to admit the patient than when they are in the lowest decile of cognitive load. The number of diagnostic orders placed is monotonically increasing in the cognitive decile, with a physician in the top decile placing 42% more diagnostic orders for the patient than when they are in the lowest decile of cognitive load, all else equal.

Using these results as a basis, we simulate a scheduling algorithm in which arriving patients are reallocated to the physician exhibiting the lowest current cognitive load, as measured by Equation (6). 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

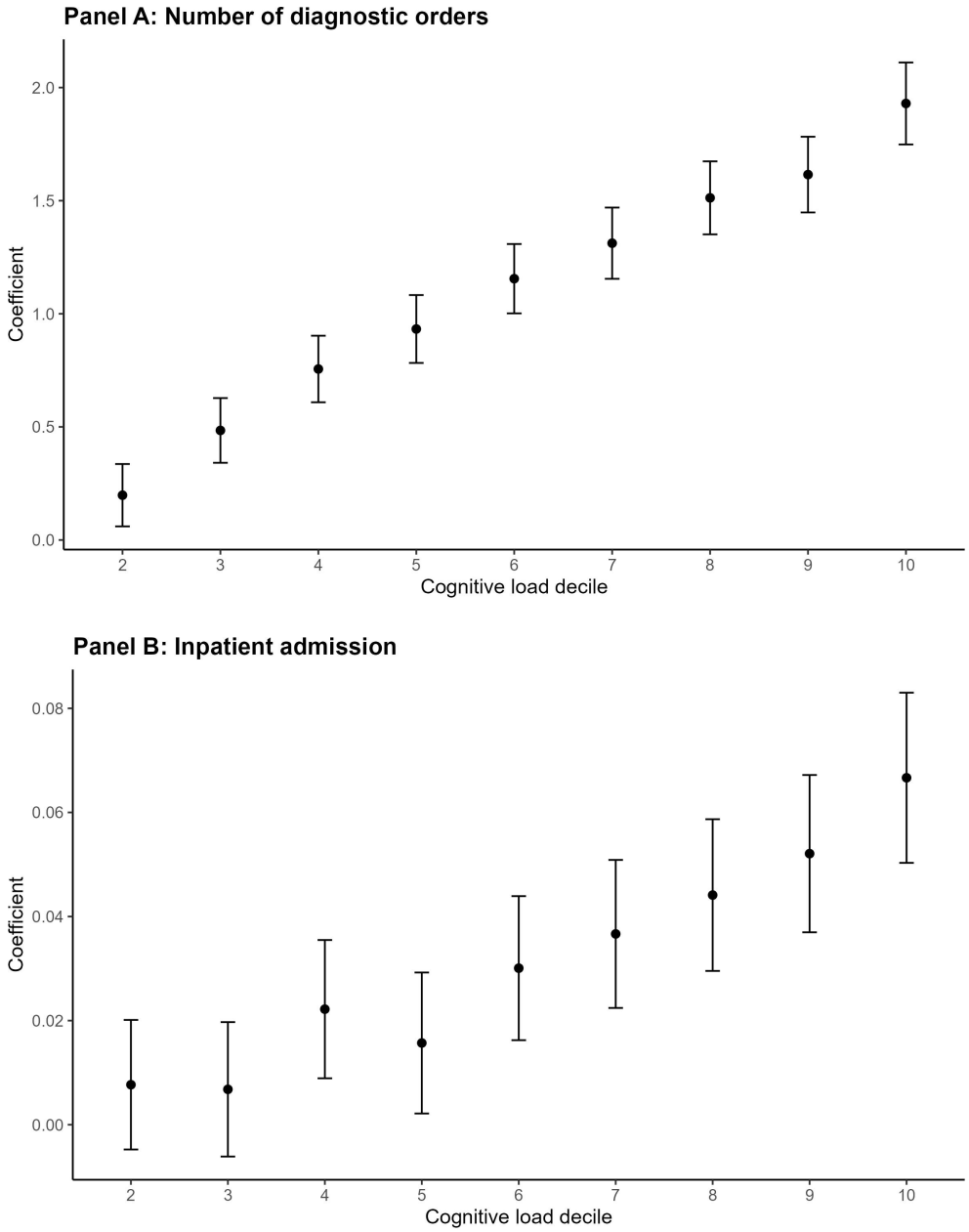


Figure 7: Estimates of the Impact of Decile of Cognitive Load on Number of DX Orders and Inpatient Admission at the Encounter

Notes: The figure plots coefficients from linear regressions of the number of diagnostic orders (Panel A) and a dummy variable whether the patient was admitted to the hospital during the encounter (Panel B) on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the encounter. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, acuity code, and physician fixed effects. The first decile is the omitted category. Estimates correspond to Columns (2) and (4) of Appendix Table G.14. Robust standard errors are used.

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.

Table 6: Patient-share Transition Matrix from Actual to Counterfactual Physician Category

	Counterfactual physician type				
	Resident Years 1/2	Resident Years 3+	Attending physician	Nurse practitioner	*Sum*
Resident Years 1/2	0.226	0.062	0.077	0.055	0.419
Resident Years 3+	0.081	0.097	0.048	0.036	0.262
Attending physician	0.047	0.022	0.028	0.019	0.145
Nurse practitioner	0.062	0.029	0.031	0.052	0.174
Sum	0.416	0.210	0.184	0.162	1.000

Notes: Each row reports the share of encounters that move from the physician category listed in the stub ("Actual") to the category in the corresponding column ("Counterfactual"). The final column shows the row sum; the final row shows column sums, and the entire table sums to 1.

Figure **??** displays the kernel density estimate of cognitive load under actual versus 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. Interestingly, the counterfactual distribution has two peaks, likely corresponding to times when the ED is busy vs. times when the ED is not busy. Still, it is notable that, even using this fairly coarse mechanism, patients under this counterfactual assignment are treated by providers with much lower cognitive load.

Table **??** presents the transition matrix of physician assignments resulting from reassignment based on cognitive load. One might be concerned that the algorithm significantly shifts assignments from one type of provider to another, conflating reductions in cognitive load with changes in provider type. Reassuringly, the table shows a stable transition matrix, with similar assignment shares within provider category in actual vs. counterfactual assignments. The off-diagnoal values are also not large, suggesting no systematic cross-substitution at a granular level.

Figure **??** illustrates the differences in the distribution of diagnostic orders under the counterfactual reassignmet relative to actual observations. This figure highlights how, as a result of the compression in the cognitive load, the order distribution also becomes more compressed and, on average, lower in terms of the mean.

Table 7 presents the average impacts of the reassignment on the key outcomes of (i) number of diagnostic orders placed and (ii) hospital admissions. Our simulation suggests that a reallocation based on cognitive load would reduce the number of diagnostic orders by roughly 0.79 (18%) per encounter. Following our earlier results, it is also likely that while reducing the number of overall orders, higher informativeness overall would result from the diagnostic orders placed. For

admissions, the reassignment reduces admissions by 7%, highlighting the potential for impacting this important decision via patient reassignment. Further analysis along these lines, with a more developed counterfactual assignment model and some additional assumptions, could assess the impact of other policies, e.g., increases in ED staffing (though accounting for quality differences would be potentially valuable, as it would also be here).

Table 7: Actual Versus Counterfactual Means

	Actual	Counterfactual
Mean number of diagnostic orders	4.552	3.760
Mean inpatient admission rate	0.238	0.222

8 Conclusion

We use novel click-stream data on physician actions in the emergency department to unpack the impacts of cognitive load on medical care. Much of the literature to date on cognitive burden and decision-making studies lab settings or smaller stakes field settings. We extend this literature with an investigation of cognitive load in a high-stakes empirical setting with experts. We observe highly detailed data on case-by-case factors that we can link to beliefs and actions, and also study a setting where there is significant and measurable exogenous variation in cognitive load. We set up a new model of the physician's diagnostic process, incorporating cognitive load and learning, and tie our empirical outcomes to comparative static predictions from the model.

We find that cognitive load impacts care in myriad ways. Holding all else equal for a given patient, higher cognitive load for a physician increases the number of diagnostic orders placed for the patient, shifts orders from more rare targeted diagnostic orders to more common diagnostic orders, and leads to order sets with lower hit rates for abnormal results. We construct a measure of diagnostic entropy to diagnostic belief refinement, and find that higher cognitive load leads to diagnostic order sets that do a worse job in refining beliefs about potential diagnoses. We show that a one standard deviation increase in cognitive load is associated with a 5% increase in inpatient hospital admission from the ED, a meaningful change in a very costly and impactful health care decision. We show that our results hold for specific chief complaints (chest pain and abdominal pain) and are robust to a variety of specification changes for (i) our construction of cognitive load measures and (ii) our definition of key outcome variables.

We see a number of exciting paths forward. Our simple counterfactual analysis studies patient reassignment accounting for cognitive load, holding a number of contextual factors fixed. There is room to extend this analysis going forward in a number of directions including, e.g., constructing a more complete dynamic assignment model and then integrating structural estimates of our con-

ceptual diagnosis model as inputs into various policy analyses leveraging that new assignment model.

Additionally, while we have a granular chatacterization of cognitive load and an individual physician's typical behavior, one could imagine incoporating (i) a richer notion of physician heterogeneity and (ii) studying interactions between that heterogeneity and cognitive load. For example, right now our empirical specifications leverage physician fixed effects to control for differences in behavior across physicians. In practice, physicians likely have different practice patterns, i.e. different exploration paths in their diagnostic processes. It will be interesting to model within-patient physician diagnostic learning paths in greater detail, and study the interaction between those distinct behavioral models and cognitive load. There are several other related interesting avenues to explore including (i) modeling entropy reduction with a fully dynamic specification moving all the way through the patient encounter and (ii) understanding how team-based production interacts with cognitive load, i.e., by smoothing the tails of negative shocks from cognitive load.

Finally, there are several directions to expand research on cognitive load in settings like ours. Researches could implement lab-in-the-field or field experiments digging more into specific cognitive load mechanisms in our context, and continue to leverage granular administrative data to enhance our understanding of how cognitive load operates. This could build closely on the recent experimental literature on cognitive load (e.g., Enke and Graeber (2023)). More broadly, with structural estimates from our conceptual diagnosis model, one could make some progress studying how artificial intelligence tools and/or certain kinds of physician training could improve care in settings with cognitive load. Of course, such studies could have broader impacts than just honing in on cognitive load: it is well known that physicians have different practice styles and that there is still much research to do investigating how altering those practice styles (e.g., to mimic top performing doctors in a granular way) can positively impact productivity (Chandra and Staiger (2007)). Above and beyond leveraging structural estimates on their own, it would be valuable to deploy artificial intelligence tools or training programs in the field using carefully crafted experiments or interventions and then leverage detailed methods and data like we use here to better understand the impacts of such programs.

References

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh, "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care," *American Economic Review*, December 2016, *106* (12), 3730–64.

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz, "Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology," Working Paper 31422, National Bureau of Economic Research July 2023.

Arrow, Kenneth J., "Uncertainty and the welfare economics of medical care," *American Economic Review*, December 1963, *53* (5), 941–973.

Badinski, Ivan, Amy Finkelstein, Matthew Gentzkow, and Peter Hull, "Geographic Variation in Healthcare Utilization: The Role of Physicians," Working Paper 31749, National Bureau of Economic Research October 2023.

Bordalo, Pedro, Nicola Gennaioli, Giacomo Lanzani, and Andrei Shleifer, "A Cognitive Theory of Reasoning and Choice," 2025. Harvard working paper.

Caplin, Andrew and Mark Dean, "Revealed preference, rational inattention, and costly information acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

Chan, David, "The Efficiency of Slacking Off: Evidence from the Emergency Department," *Econometrica*, 2018, *86* (3), 997–1030.

Chan, David C., "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy*, 2016, *124* (3), 734–770.

Chan, David C, Matthew Gentzkow, and Chuan Yu, "Selection with Variation in Diagnostic Skill: Evidence from Radiologists*," *The Quarterly Journal of Economics*, 01 2022, *137* (2), 729–783.

Chandra, Amitabh and Douglas O. Staiger, "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy*, 2007, *115* (1), 103–140.

_ , David Cutler, and Zirui Song, "Chapter Six - Who Ordered That? The Economics of Treatment Choices in Medical Care," in Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds., *Handbook of Health Economics*, Vol. 2 of *Handbook of Health Economics*, Elsevier, 2011, pp. 397–432.

Chodick, Gabriel, Yoav Goldstein, Ity Shurtz, and Dan Zeltzer, "Challenging Encounters and within-Physician Practice Variability," *The Review of Economics and Statistics*, 05 2025, pp. 1–9.

Currie, Janet and W. Bentley MacLeod, "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians," *Journal of Labor Economics*, 2017, *35* (1), 1–43.

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg, "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," *American Economic Journal: Economic Policy*, February 2019, *11* (1), 192–221.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso, "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences*, 2011, *108* (17), 6889–6892.

Dean, Mark and Nathaniel Neligh, "Experimental tests of rational inattention," *Journal of Political Economy*, 2023, *131* (12), 3415–3461.

Enke, Benjamin and Thomas Graeber, "Cognitive Uncertainty*," *The Quarterly Journal of Economics*, 05 2023, *138* (4), 2021–2067.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, "Sources of Geographic Variation in Health Care: Evidence From Patient Migration*," *The Quarterly Journal of Economics*, 07 2016, *131* (4), 1681–1726.

Gabaix, Xavier, "Chapter 4 - Behavioral inattention," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics - Foundations and Applications 2*, Vol. 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, North-Holland, 2019, pp. 261–343.

Hsu, Chao-Chun, Ziad Obermeyer, and Chenhao Tan, "Clinical Notes Reveal Physician Fatigue," 2023.

Kannampalil, Thomas and Julia Adler-Millstein, "Using electronic health record audit log data for research: insights from early efforts," *Journal of the Medical Informatics Association*, December 2022, *30* (1), 167–171.

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach, "Do financial concerns make workers less productive?," *The Quarterly Journal of Economics*, 2025, *140* (1), 635–689.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions*," *The Quarterly Journal of Economics*, 08 2017, *133* (1), 237–293.

Matějka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 2015, *105* (1), 272–298.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Review," *Journal of Economic Literature*, March 2023, *61* (1), 226–73.

McGuire, Thomas G., "Physician agency," in A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics*, 1 ed., Vol. 1, Elsevier, 2000, chapter 09, pp. 461–536.

Mullainathan, Sendhil and Ziad Obermeyer, "A Machine Learning Approach to Low-Value Health Care: Wasted Tests, Missed Heart Attacks and Mis-Predictions," Working Paper 26168, National Bureau of Economic Research August 2019.

Oprea, Ryan, "Decisions under Risk Are Decisions under Complexity," *American Economic Review*, December 2024, *114* (12), 3789–3811.

Silver, David, "Haste or Waste? Peer Pressure and Productivity in the Emergency Department," *The Review of Economic Studies*, 08 2020, *88* (3), 1385–1417.

Skinner, Jonathan, "Chapter Two - Causes and Consequences of Regional Variations in Health Care11This chapter was written for the Handbook of Health Economics (Vol. 2). My greatest debt is to John E. Wennberg for introducing me to the study of regional variations. I am also grateful to Handbook authors Elliott Fisher, Joseph Newhouse, Douglas Staiger, Amitabh Chandra, and especially Mark Pauly for insightful comments, and to the National Institute on Aging (PO1 AG19783) for financial support.," in Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds., *Handbook of Health Economics*, Vol. 2 of *Handbook of Health Economics*, Elsevier, 2011, pp. 45–93.

Song, Yunjie, Jonathan Skinner, Julie Byn6um, Jason Sutherland, John Wennberg, and Elliot Fisher, "Regional Variations in Diagnostic Practices," *New England Journal of Medicine*, 2010, *363*, 45–53.

Wellsjo, Alex Steiny, "Simple Actions, Complex Habits: Lessons from Hospital Hand Hygiene," 2025. UCSD working paper.

APPENDIX

A Proofs for Section 2

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. □

Proof of Proposition 1. Number of mental signals n^M is that largest $n > 0$ such that $n < \mu^E - L$,[25] 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 decreasing in L. Conditional on a chosen n^M the objective (3) then takes a form $(-\alpha 2^{-n^E} - \mu^E n^E)$, where $\alpha = 2/D^{n^M+1}$ is decreasing in n^M. Higher L, and thus higher α decreases the objective, but less so for higher n^E since $\partial^2(-\alpha 2^{-n^E} - \mu^E n^E)/\partial\alpha\partial n^E = 2^{-n^E}\log 2 > 0$. First order condition for n^E is thus satisfied at a weakly higher n^E. Higher L increases n^E. □

[25] Discrete nature of the information acquisition problem could generate non-uniqueness. For simplicity of exposition here we assume a tie-breaker such that if there is more than one pair of $\{n^M, n^E\}$ delivering the same maximal objective, the physicians chooses among those with the lowest n^M. Weak monotonicities, however, hold without it as well.

B Details on Emergency Department Environment

Patients arrive either by ambulance or by walk-in to the front-desk station. A registrar collects demographic identifiers and opens an electronic encounter in Epic. This "rapid registration" usually takes less than 2 minutes and does not influence subsequent triage priority.

A licensed triage nurse then performs a focused assessment in a dedicated triage bay (or at the bedside for direct-to-room arrivals). Vital signs, symptom onset, and brief history are obtained and recorded, together with any required isolation flags. Using the Emergency Severity Index (ESI, version 5) algorithm the nurse assigns an acuity code from 1 (immediate life-saving intervention) to 5 (no resources anticipated). The nurse simultaneously selects a structured chief-complaint label from Epic's menu (e.g., "chest pain", "abdominal pain") or enters free text; this label is the field displayed in Appendix Figure **??**.

Once triage is complete, patients are either taken straight to an available bed ("direct-bed") or returned to the waiting room. Their queue position is managed by the charge nurse but displayed to all clinicians on the live computerized tracking board (shown in Appendix Figure **??**). The board lists, in order of arrival within each acuity tier, the chief complaint, the elapsed waiting time and any preliminary nursing orders such as point-of-care tests.

Physicians pick up cases sequentially ("next-up") from this board. At the moment of selection they observe only the chief complaint and acuity code – as reflected in our control variables.

This workflow means that i) initial assignment of acuity level and chief complaint is nurse-driven and completed before any physician becomes involved, and ii) within each time block the set of patients visible to a given doctor is governed almost entirely by arrival sequence and bed availability rather than by patient characteristics beyond acuity and complaint. These features motivate our quasi-random assignment assumption and the control strategy described in Section 3.1.

C Predicting Patient Complexity

C.1 Constructing the Patient Complexity Score

Our patient-complexity score c_i is designed to proxy the *expected* time and cognitive effort a physician must invest in encounter i. The construction proceeds in four steps.

1. Define three workload targets For every encounter we extract:

i) y_{1i}: **Total number of diagnostic and therapeutic orders** placed in the ED (medications, labs, imaging, procedures,...)

ii) y_{2i}: **Total note-edit time** in seconds, transformed as $\log(1 + y_{2i})$ to down-weight the right tail.

iii) y_{3i}: **Number of specialty consultations** initiated in the ED.

2. Assemble encounter-level predictors $\mathbf{X}_i$ All predictors are observable at, or immediately after, triage:

- *Demographics and comorbidity:* age-bin dummies (18–29, 30–44, $\ldots, \geq$ 90), gender indicator, Charlson comorbidity index (continuous), and an indicator for interpreter need.
- *Chief complaint:* three mutually exclusive sets of dummies – Tier 1 (15 body-system buckets), Tier 2 (72 sub-systems), and the detailed chief-complaint code (732 categories).
- *Clinical acuity and vitals:* ESI acuity code (1–5), six binary "abnormal vital" flags (systolic/diastolic BP, pulse, respiration, temperature, O_2 saturation).
- *Utilization history:* indicators for any ED visit or inpatient admission in the preceding 60 days and for means of arrival (ambulance, police, walk-in, ...)
- *Fixed effects:* 24 hour-of-day and 7 day-of-week dummies.

Continuous variables are z-scored, categorical variables are one-hot-encoded with the first level omitted.

3. Predict each workload target with LASSO regression For $m \in \{1, 2, 3\}$ we regress y_{mi} on $\mathbf{X}_i$ using a five-fold cross-validated LASSO:

$$\hat{\beta}_m \;=\; \arg\min_{\beta} \; \frac{1}{N} \sum_i (y_{mi} - \mathbf{X}_i'\beta)^2 + \lambda \|\beta\|_1, \tag{12}$$

where the penalty λ is chosen by the minimum average cross-validated MSE (Appendix Figure C.1). We then subtract the fitted contributions of the hour- and day-dummies so that the resulting predicted values, $\hat{y}_{mi}^{\text{FE-adj}}$, vary only with patient covariates.

4. Create a single complexity index Each FE-adjusted prediction is standardized:

$$\tilde{y}_{mi} = (\hat{y}_{mi}^{\text{FE-adj}} - \mu_m)/\sigma_m \tag{13}$$

The composite workload proxy is their simple average:

$$\tilde{c}_i = \tfrac{1}{3} \sum_{m=1}^{3} \tilde{y}_{mi}. \tag{14}$$

We standardize once more and shift by the absolute minimum so that

$$c_i \;=\; \frac{\tilde{c}_i - \mu_c}{\sigma_c} \;+\; \left|\min_j \tfrac{\tilde{c}_j - \mu_c}{\sigma_c}\right| \;\;>\; 0. \tag{15}$$

The distribution is right-skewed with a primary mode near 3 (Appendix Figure 4).

C.2 Alternative LASSO Specifications

We estimate six variants of the prediction exercise to verify the robustness of c_i:

Chief-complaint covariates	Full sample	Two-half split
Tier 2 only	✓	✓
Detailed code only	✓	✓
Tier 1 + Tier 2 + detailed	✓	✓

Two-half sample split: To eliminate any possibility of over-fitting "leaking" into the main analysis, we randomly assign encounters to halves A and B. A model trained on A predicts complexity for B and *vice-versa*. Each observation is therefore scored by a model that never used its own outcome data.

Chief-complaint granularity: Using only Tier 2, only the detailed code, or all three levels in the design matrix tests whether the predictive content of chief complaint depends on resolution.

Across all six variants the within-sample R^2 values differ by < 0.01 (Appendix Table C.4), and the pairwise correlations of the resulting complexity indices exceed 0.96 (same table). We adopt the *full-sample, Tier1 + Tier2 + detailed* specification for the baseline results for the sake of simplicity.

C.3 Summary Statistics of 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). 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.

Table C.1: Summary Statistics for LASSO Outcomes and Continuous Predictors

	Min	P10	P25	Mean	Median	P75	P90	Max
Outcomes								
Number of all orders in ED	0	1	4	16.737	11	23	41	209
Number of consults	0	0	0	0.232	0	0	1	13
Log(1 + Total ED note edit time)	0.000	5.787	6.268	6.809	6.834	7.515	8.118	10.041
Continuous predictors								
Charlson comorbidity index	0	0	0	0.617	0	0	2	16
Number of abnormal vital signs	0	0	0	0.565	0	1	2	6

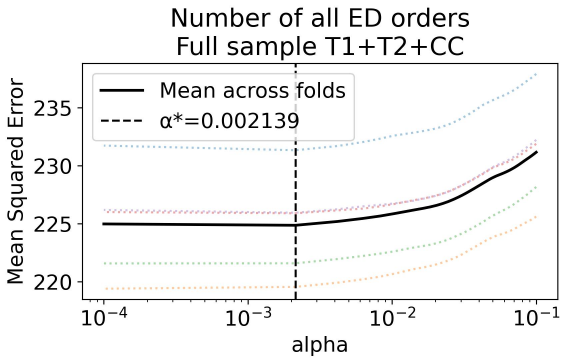
Notes: The table shows summary statistics for the three outcomes and the two continuous predictors of the LASSO complexity prediction model.

Table C.2: Summary Statistics for Categorical LASSO Predictors

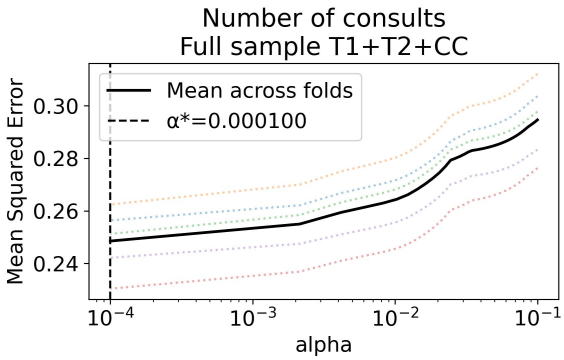
	Share		Share
Demographics		**Chief complaint Tier 1** (15 total categories)	
Gender: Male	0.510	Neurological	0.147
Interpreter needed	0.102	GI/Abdominal	0.143
		Musculoskeletal	0.135
Health status		Other	0.099
Admitted within last 60 days	0.432	Cardiovascular	0.085
ED visit within last 60 days	0.217	Respiratory	0.072
		Trauma	0.068
Vitals		HEENT	0.049
Abnormal systolic blood pressure	0.099	Psychiatric	0.049
Abnormal diastolic blood pressure	0.316	GU	0.045
Abnormal temperature	0.004	**Chief complaint Tier 2** (72 total categories)	
Abnormal pulse oximetry	0.067	MSK pain	0.110
Abnormal Respiration	0.052	Abdominal pain	0.098
Abnormal pulse	0.028	Chest pain	0.062
		Respiratory problem	0.057
Acuity level		Neurological symptoms	0.038
Immediate	0.004	Psychiatric problem	0.037
Emergent	0.194	Vomiting/diarrhea	0.030
Urgent	0.599	Fall	0.029
Less Urgent	0.180	Headache	0.028
Non-Urgent	0.019	Altered Mental Status	0.027
Missing	0.003	**Chief complaint** (732 total categories)	
		Abdominal Pain	0.096
Age bin		Chest Pain	0.062
0-17	0.011	Shortness of Breath	0.053
18-29	0.186	Fall	0.029
30-39	0.162	Altered Mental Status	0.027
40-49	0.142	Headache	0.026
50-59	0.169	Back Pain	0.025
60-69	0.160	Dizziness	0.023
70-79	0.106	Psychiatric Evaluation	0.022
80+	0.064	Weakness	0.022
N = 84,214			

Notes: The table shows the category shares of all categorical predictors used in the LASSO model to predict patient complexity. For the chief complaint variables (Tier 1, Tier 2, and detailed), only the ten most common categories are shown.

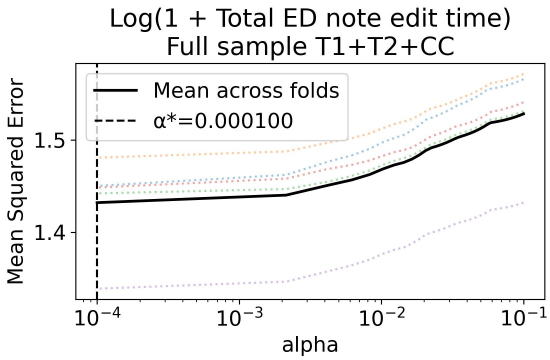
C.4 Evaluating Lasso Predictions



Panel A: Number of all ED orders



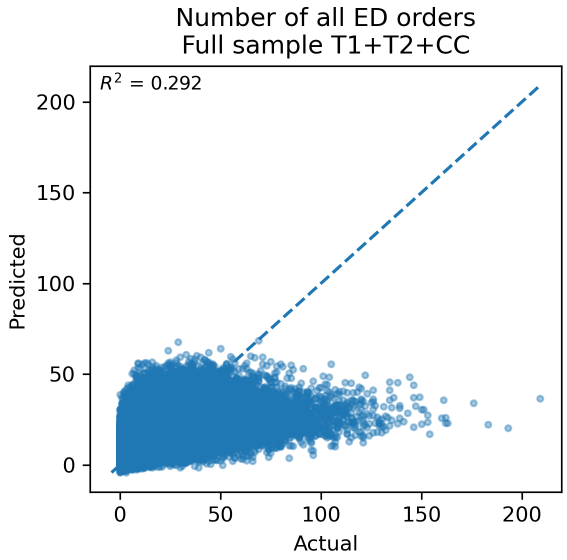
Panel B: Number of consults



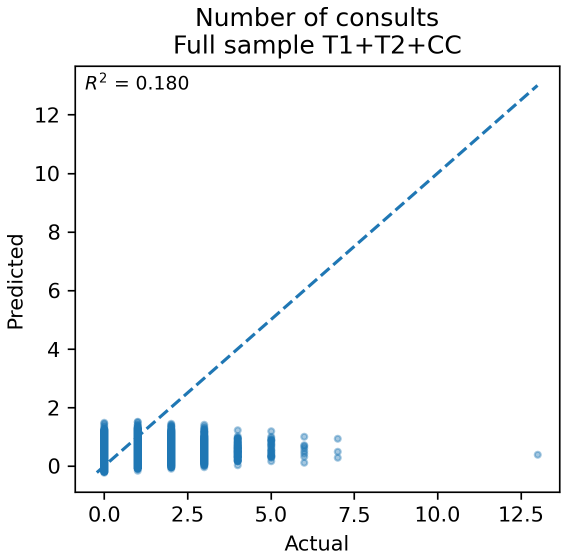
Panel C: Log(1 + Total ED note edit time)

Figure C.1: Five-fold LASSO cross-validation results

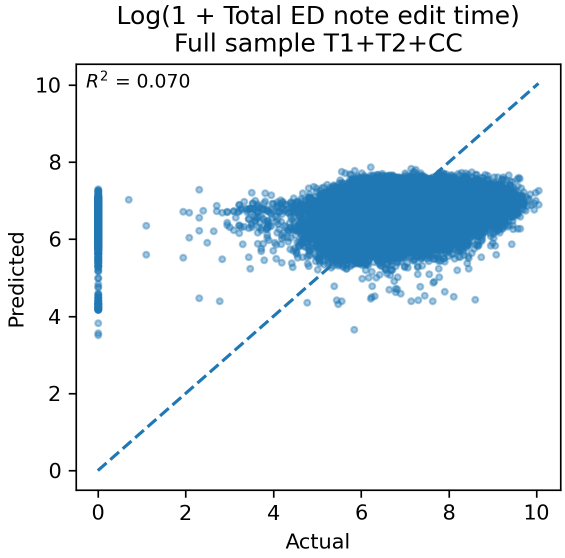
Notes: All results are for the full-sample specification using all three chief complaint categorical variables as predictors: Tier 1 (T1), Tier 2 (T2), and detailed (CC) chief complaint. Dotted lines show fold-specific mean-squared-error (MSE) trajectories across the tested regularization parameter (α) grid; the solid black line is the average across folds. The vertical dashed line marks the α value chosen by cross-validation.



Panel A: Number of all ED orders



Panel B: Number of consults



Panel C: Log(1 + Total ED note edit time)

Figure C.2: Comparison of model predictions to observed values.

Notes: All results are for the full-sample specification using all three chief complaint categorical variables as predictors: Tier 1 (T1), Tier 2 (T2), and detailed (CC) chief complaint. Each panel plots encounter-level predictions (vertical axis) against actual outcomes (horizontal axis); the 45° dashed line indicates perfect prediction. The within-sample R^2 for each outcome is reported in the upper-left corner of its panel.

C.5 Comparing Different Lasso Prediction Models

Table C.3: Goodness-of-Fit (R^2) for Patient Complexity Measures

	Number of all ED orders	Log(Note edit time)	Number of consults
Full sample T1 + T2 + CC	0.292	0.070	0.180
Sample split T1 + T2 + CC	0.290	0.068	0.172
Full sample CC	0.292	0.069	0.180
Sample split CC	0.290	0.067	0.177
Full sample T2	0.286	0.066	0.164
Sample split T2	0.285	0.064	0.163

Notes: Entries show within-sample R^2 for each LASSO model. "Sample split" fits on one half of the data and is evaluated on the other; "Full sample" uses the entire dataset. Chief-complaint detail enters the LASSO as Tier 2 indicators ("T2"), the detailed chief-complaint code ("CC"), or all three levels combined ("T1 + T2 + CC").

Table C.4: Correlation Matrix for Patient Complexity Measures

	T2 (split)	T2 (full)	CC (split)	CC (full)	T1+T2+CC (split)	T1+T2+CC (full)
T2 (split)	1.000	0.998	0.983	0.982	0.989	0.984
T2 (full)	0.998	1.000	0.982	0.984	0.987	0.986
CC (split)	0.983	0.982	1.000	0.996	0.995	0.995
CC (full)	0.982	0.984	0.996	1.000	0.993	0.999
T1 + T2 + CC (split)	0.989	0.987	0.995	0.993	1.000	0.994
T1 + T2 + CC (full)	0.984	0.986	0.995	0.999	0.994	1.000

Notes: Each entry shows the Pearson correlation between two encounter-level complexity indices. All indices are standardized, shifted versions of the underlying composite score described in Section 2.3. "(split)" refers to "Sample split": these models are fit on one half of the data and evaluated on the other; "(full)" refers to "Full sample": these models are estimated on the entire dataset. Chief-complaint information enters the LASSO as either Tier 2 indicators ("T2"), the detailed chief-complaint code ("CC"), or all three levels combined ("T1 + T2 + CC").

Table C.5: LASSO Feature Importance: Number of all ED orders

Variable	Coefficient
Chief complaint: Pneumonia	12.469
Acuity code: Non-Urgent	-12.257
Chief complaint: Diabetic Ketoacidosis	11.181
Chief complaint: Diplopia	11.072
Acuity code: Less Urgent	-10.681
Chief complaint: Multiple Sclerosis	9.984
Chief complaint Tier 1: Endocrine	9.472
Chief complaint: Sickle Cell Pain Crisis	8.860
Chief complaint: Meningitis	8.220
Chief complaint: Spasms	8.006
Chief complaint: Hand Problem	6.880
Chief complaint: Arm Swelling	6.779
Chief complaint: Gait Problem	6.694
Chief complaint: Hallucinations	-6.347
Chief complaint: Hypertension	-6.250
Chief complaint: Extremity Weakness	5.749
Acuity code: Missing	-5.500
Means of arrival: Transfer - System Arrival	-5.103
Chief complaint Tier 1: Constitutional	4.728
Chief complaint Tier 2: Chest pain	4.364
Chief complaint: Brain Tumor	-4.241
Chief complaint: Cellulitis/ Skin Infection	4.234
Chief complaint: Sciatica	-4.233
Chief complaint: Shortness of Breath	4.192
Chief complaint Tier 2: Flank pain	4.172

Notes: This table shows the 25 most important predictors for the outcome "Number of all ED orders" as measured by the absolute magnitude of the LASSO coefficient.

Table C.6: LASSO Feature Importance: Log(1 + Total ED note edit time)

Variable	Coefficient
Chief complaint: Other	-1.015
Acuity code: Missing	-1.004
Means of arrival: Unknown	-0.823
Acuity code: Non-Urgent	-0.579
Chief complaint Tier 2: Social problems	-0.414
Chief complaint: Drug / Alcohol Assessment	-0.327
Chief complaint: Alcohol Intoxication	-0.282
Chief complaint: Hallucinations	-0.277
Chief complaint: Asthma	-0.243
Acuity code: Emergent	0.225
Chief complaint Tier 1: Drug use	-0.211
Chief complaint: Leg Injury	-0.197
Chief complaint Tier 2: Psychiatric problem	-0.184
ED visit within last 60 days: 1	-0.170
Acuity code: Urgent	0.166
Age bin: 30-39	-0.164
Acuity code: Less Urgent	-0.163
Chief complaint: Shoulder Injury	-0.160
Chief complaint Tier 2: Blood disorders	0.156
Chief complaint Tier 2: Weight loss	0.136
Chief complaint: Rash	0.135
Chief complaint: Psychiatric Evaluation	0.134
Age bin: 40-49	-0.129
Chief complaint: Cerebrovascular Accident	-0.125
Age bin: 18-29	-0.124

Notes: This table shows the 25 most important predictors for the outcome "Log(1 + Total ED note edit time)" as measured by the absolute magnitude of the LASSO coefficient.

Table C.7: LASSO Feature Importance: Number of Consults

Variable	Coefficient
Chief complaint: Multiple Sclerosis	0.590
Chief complaint Tier 2: Suicidal	0.532
Chief complaint: Airway Obstruction	0.481
Chief complaint: Urinary Incontinence	0.442
Chief complaint: Diplopia	0.433
Chief complaint: Zoster keratitis	0.430
Chief complaint: Circulatory Problem	0.421
Chief complaint: Manic Behavior	0.384
Chief complaint: Depression	0.369
Chief complaint: Psychiatric Evaluation	0.355
Chief complaint: Aggressive Behavior	0.352
Chief complaint: Foot Ulcer	0.344
Chief complaint: Skin Lesion	0.330
Chief complaint: Shingles/ Herpes Zoster	0.284
Chief complaint: Sciatica	-0.274
Chief complaint: Loss of Vision	0.270
Chief complaint: Gait Problem	0.266
Chief complaint: Stress	-0.266
Chief complaint: Floaters	0.255
Chief complaint: GI Problem	0.251
Chief complaint: Anal Pain	0.247
Means of arrival: Police	0.240
Acuity code: Non-Urgent	-0.231
Chief complaint: Panic Attack	-0.228
Chief complaint Tier 2: Eye problem	0.227

Notes: This table shows the 25 most important predictors for the outcome "Number of consults" as measured by the absolute magnitude of the LASSO coefficient.

D Additional Descriptive Figures and Tables

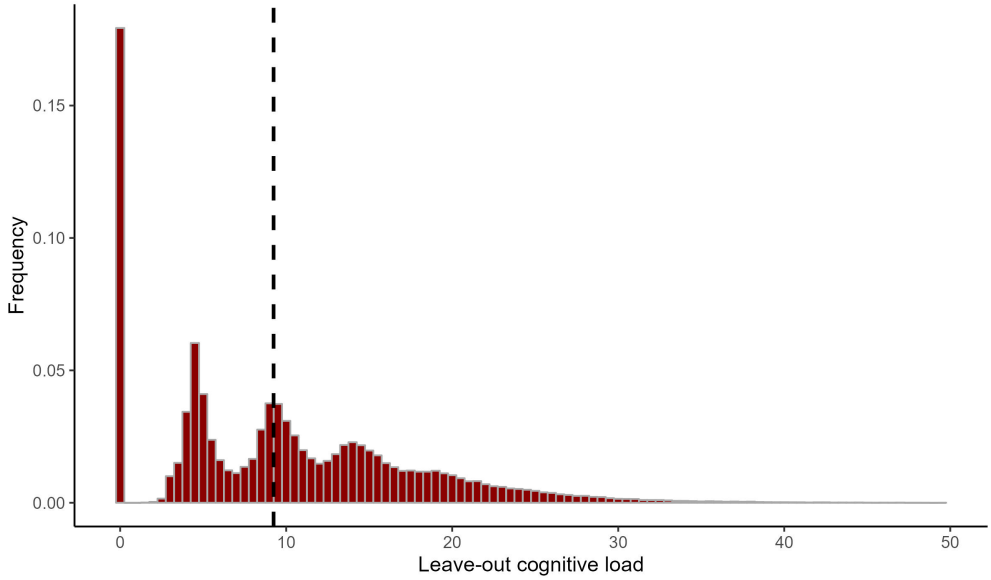


Figure D.1: Distribution of Leave-Out Cognitive Load

Notes: The figure shows the relative frequencies of the leave-out cognitive load measure with a binwidth of 0.5. Each observation is the cognitive load at one *encounter-action*. The vertical dashed line indicates the median.

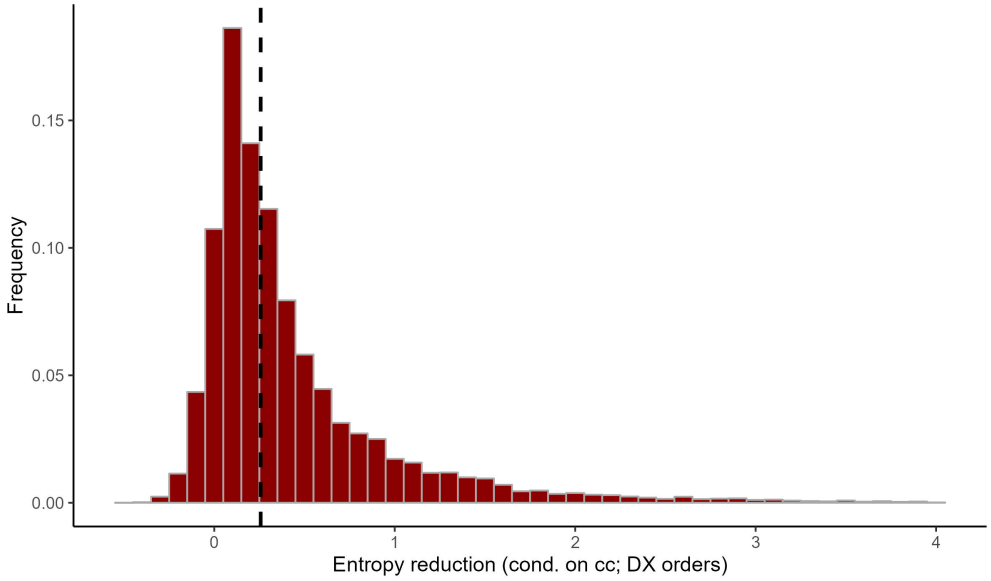


Figure D.2: Distribution of Entropy Reduction in First Order Batch

Notes: The figure shows the relative frequencies of the entropy reduction in the first-order batch with a binwidth of 0.1. Each observation is the entropy reduction for one *encounter*. The vertical dashed line indicates the median.

Table D.1: Shift summary statistics by physician type

	Min	Mean	Median	Max
Residents years 1/2				
Number of shifts	1	20.279	12	142
Average shift length (hours)	7.305	11.786	10.478	28.801
Average patients per shift	1.000	3.505	3.467	9.571
Residents years 3+				
Number of shifts	1	15.642	2	101
Average shift length (hours)	5.165	13.418	12.258	29.178
Average patients per shift	1.000	3.236	1.000	11.889
Attending physicians				
Number of shifts	1	71.897	13	514
Average shift length (hours)	5.387	11.270	10.734	24.794
Average patients per shift	1.000	1.479	1.217	3.431
Nurse practitioner				
Number of shifts	1	92.733	54	412
Average shift length (hours)	7.251	11.137	10.899	15.955
Average patients per shift	1.000	3.487	3.437	8.379

Notes: Each row reports summary statistics for shifts at the *physician* level for the indicated provider category.

Table D.2: 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

Notes: The table shows the 'display names' of the 25 most frequently placed diagnostic ED orders, ranked by their share of all diagnostic orders.

Table D.3: 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

Notes: The table shows the 'display names' of the 25 most frequently placed medication ED orders, ranked by their share of all medication orders.

Table D.4: 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

Notes: The table shows the 25 most common chief complaints (out of 732 distinct chief complaints).

Table D.5: 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

Notes: The table shows the 25 most common diagnoses.

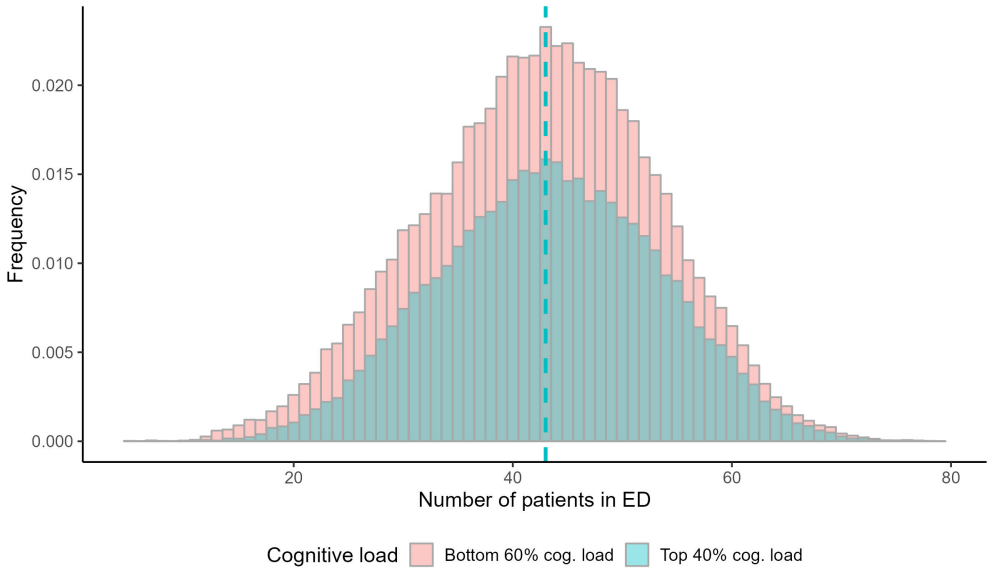


Figure D.3: Distribution of Number of Patients in ED

Notes: The figure shows the distribution of the number of patients in the ED, separately for the top two quintile (blue) and the bottom three quintiles (red). The vertical dashed line shows the medians of both distributions, which overlap exactly.

E Patient-Level Examples of Diagnostic Entropy Reduction

This section illustrates how entropy reduction varies across two abdominal-pain encounters, detailing the diagnostic orders and the calculation procedure introduced in Section 6.1 of the main text.

Patient Example 1

This example follows the entropy-reduction procedure introduced in Section 6.1 and Equations 10–11 of the main text.

Clinical context: An adult patient presents with *abdominal pain*. Four point-of-care (POC) or laboratory urine studies are ordered within 30 minutes. The encounter's final diagnosis is *urinary-tract infection* (UTI).

Order 1: *POC Urinalysis Dipstick.* Screens urine for glucose, protein, blood, leukocyte esterase, nitrites, ketones and more, offering an immediate infection- or renal-injury signal.

Order 2: *POC Urine Pregnancy Test.* Detects human chorionic gonadotropin (hCG) in minutes, ruling out pregnancy before imaging or medication that is teratogenic.

Order 3: *Laboratory Urinalysis with Reflex-to-Culture.* Microscopic sediment exam; automatically cultures if pyuria or bacteriuria exceeds preset thresholds.

Order 4: *Urine Culture from Screening Urinalysis.* Provides definitive pathogen identification and antimicrobial sensitivities for targeted therapy.

Order-specific diagnosis distributions: Figure E.1 displays four separate histograms - one per order - showing how often each final diagnosis occurs in encounters where that order was placed for abdominal-pain presentations.

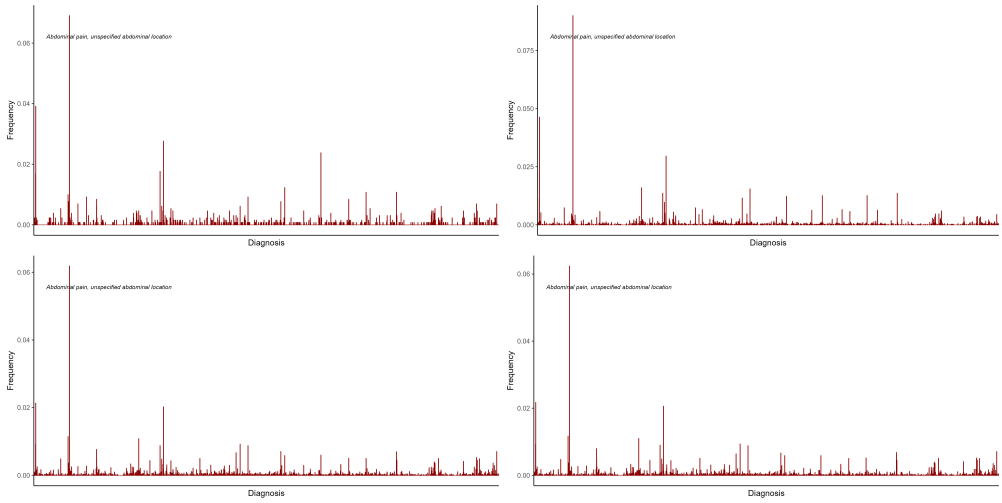


Figure E.1: Diagnosis Distributions Implied by Orders 1–4

Notes: Each panel shows the empirical frequency of final diagnoses associated with an *individual* order, conditional on chief complaint *abdominal pain*. The vertical line marks the diagnosis most frequently linked to that specific order.

Aggregated diagnosis distribution: Figure E.2 *sums* the frequencies from Figure E.1 across all four orders, normalizes them to sum to one, and visualizes the aggregate probability mass function implied by the entire order set.

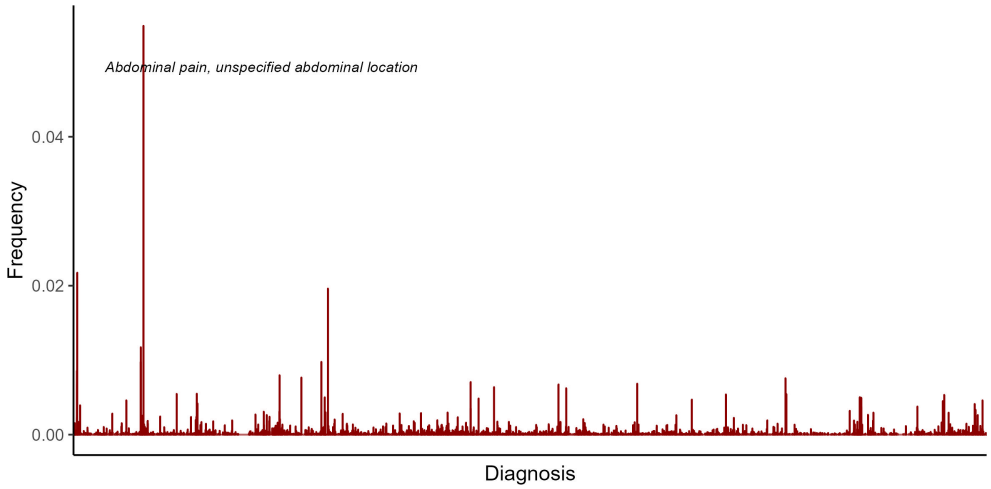


Figure E.2: Aggregated Diagnosis Distribution Across Orders 1–4

Notes: Empirical frequencies from Figure E.1 are summed across the four orders and renormalized to sum to one. The resulting distribution reflects the combined diagnostic information of the full order set. The vertical line indicates the diagnosis with the highest aggregated probability.

Entropy calculation:

1. The empirical entropy of final diagnoses for the chief complaint *abdominal pain* is

$$H(\text{Abdominal Pain}) = 6.52.$$

   A higher entropy value indicates a broader, more uncertain distribution over possible diagnoses.

2. Aggregating the marginal diagnosis distributions implied by Orders 1–4 and calculating the entropy following Equation 10 yields:

$$H(\text{Order Set} \mid \text{Abdominal Pain}) = 5.82.$$

   A *lower* entropy here reflects that the order set places probability mass on a broader subset of diagnoses.

3. Following Equation 11, entropy reduction for this encounter is

$$E_1 = H(\text{Abdominal Pain}) - H(\text{Order Set} \mid \text{Abdominal Pain}) = 0.70.$$

   A *larger* entropy reduction signifies that the physician's diagnostic orders have sharpened the implied distribution which indicates more precise diagnostic reasoning for this patient.

Patient Example 2

Clinical context: A second adult patient presents with *abdominal pain*. Four diagnostic studies—two urine-based and two blood-based—are ordered within 30 minutes. The encounter's final diagnosis is again *urinary-tract infection* (UTI).

Order 1: *Urinalysis with Microscopy.* Full dipstick plus microscopic sediment examination (cells, casts, crystals) performed up-front.

Order 2: *Complete Blood Count with Differential.* Evaluates leukocytosis or left shift that could suggest systemic infection or alternate intra-abdominal pathology.

Order 3: *Basic Metabolic Panel.* Measures electrolytes, renal function and acid–base status, guiding imaging contrast use or antibiotic dosing.

Order 4: *Venous Blood Gas with Lactate.* Provides pH, pCO_2, bicarbonate and lactate; elevated lactate flags sepsis or tissue hypoperfusion.

Order-specific diagnosis distributions: Figure E.3 displays four separate histograms – one per order – showing how often each final diagnosis occurs in encounters where that order was placed for abdominal-pain presentations.

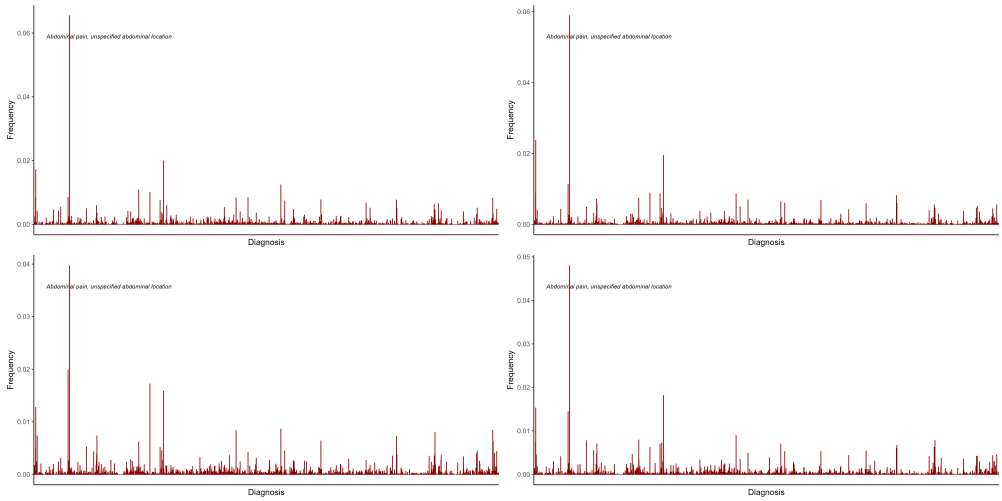


Figure E.3: Diagnosis Distributions Implied by Orders 1–4

Notes: Each panel shows the empirical frequency of final diagnoses associated with an *individual* order, conditional on chief complaint *abdominal pain*. The vertical line marks the diagnosis most frequently linked to that specific order.

Aggregated diagnosis distribution: Figure E.4 *sums* the frequencies from Figure E.3 across all four orders, normalizes them to sum to one, and visualizes the aggregate probability mass function implied by the entire order set.

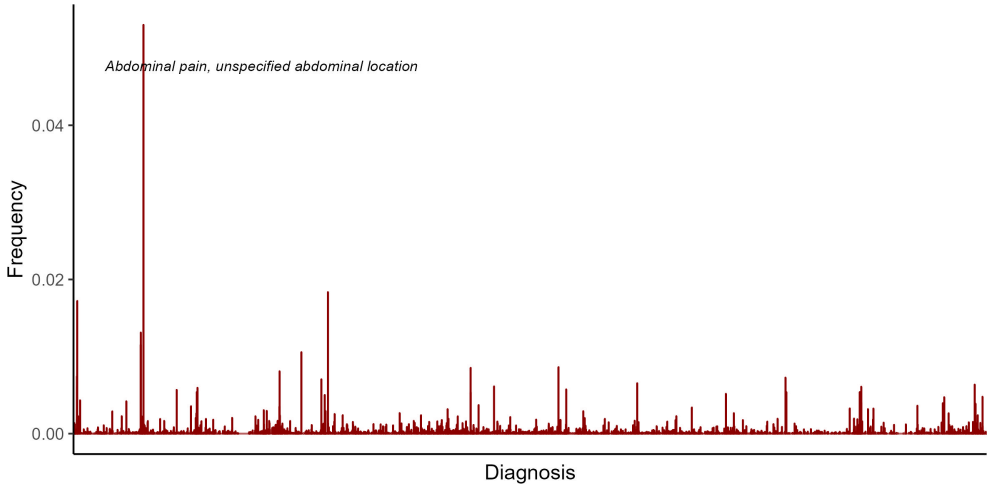


Figure E.4: Aggregated Diagnosis Distribution Across Orders 1–4

Notes: Empirical frequencies from Figure E.3 are summed across the four orders and re-normalized to sum to one. The resulting distribution reflects the combined diagnostic information of the full order set. The vertical line indicates the diagnosis with the highest aggregated probability.

Entropy calculation:

1. The empirical entropy of final diagnoses for the chief complaint *abdominal pain* is

$$H(\text{Abdominal Pain}) = 6.52.$$

   A higher entropy value indicates a broader, more uncertain distribution over possible diagnoses.

2. Aggregating the marginal diagnosis distributions implied by Orders 1–4 and calculating the entropy following Equation 10 yields:

$$H(\text{Order Set} \mid \text{Abdominal Pain}) = 6.44.$$

   A *lower* entropy here reflects that the order set places probability mass on a broader subset of diagnoses.

3. Following Equation 11, entropy reduction for this encounter is

$$E_2 = H(\text{Abdominal Pain}) - H(\text{Order Set} \mid \text{Abdominal Pain}) = 0.08.$$

   A *larger* entropy reduction signifies that the physician's diagnostic orders have sharpened the implied distribution which indicates more precise diagnostic reasoning for this patient. Conversely, the relatively small reduction here suggests a broader, less discriminating work-up.

F Alternative Regression Specifications: Encounter-Action Level

F.1 Alternative Functional Forms

F.1.1 Cognitive Load Quintiles

Table F.1: OLS Estimates of the Impact of Cognitive Load on Orders: Cognitive Load Quintiles

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Q5 cog. load	0.131***	0.097***	0.010	−0.006	0.003**	0.003**
	(0.014)	(0.015)	(0.006)	(0.007)	(0.001)	(0.002)
Q4 cog. load	0.039***	0.024*	0.005	−0.006	0.001	0.002
	(0.013)	(0.013)	(0.006)	(0.006)	(0.001)	(0.001)
Q2 cog. load	−0.044***	−0.030**	−0.014**	−0.005	−0.004***	−0.005***
	(0.013)	(0.013)	(0.006)	(0.006)	(0.001)	(0.001)
Q1 cog. load	−0.033**	−0.009	−0.018***	−0.017**	0.007***	0.006***
	(0.014)	(0.015)	(0.007)	(0.007)	(0.001)	(0.002)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.30	1.30	0.66	0.66	0.06	0.06
N	266,625	266,625	266,625	266,625	266,625	266,625
R^2	0.043	0.052	0.024	0.057	0.012	0.017
Adjusted R^2	0.040	0.047	0.021	0.052	0.010	0.012

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of orders that yield an abnormal result on quintiles of the leave-out measure of cognitive load. The unit of observation is the *encounter–event* level, where an *action* either 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), (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%).

Table F.2: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking: Cognitive Load Quintiles

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent) editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Q5 cog. load	0.149***	0.117***	−0.017***	−0.020***	−0.325***	−0.113***
	(0.013)	(0.013)	(0.006)	(0.006)	(0.012)	(0.012)
Q4 cog. load	0.051***	0.038***	−0.012**	−0.014***	−0.135***	−0.052***
	(0.011)	(0.012)	(0.005)	(0.005)	(0.012)	(0.011)
Q2 cog. load	−0.047***	−0.036***	0.004	0.005	0.131***	0.065***
	(0.011)	(0.011)	(0.005)	(0.006)	(0.012)	(0.011)
Q1 cog. load	−0.076***	−0.047***	0.043***	0.038***	0.148***	0.103***
	(0.012)	(0.013)	(0.006)	(0.006)	(0.014)	(0.013)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.046	0.054	0.022	0.032	0.023	0.112
Adjusted R^2	0.043	0.049	0.019	0.027	0.020	0.107

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 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%).

Table F.3: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision: Cognitive Load Quintiles

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Q5 cog. load	-0.017^{**}	-0.016^{*}	-0.016^{*}	-0.008
	(0.008)	(0.008)	(0.008)	(0.007)
Q4 cog. load	-0.007	-0.007	-0.004	-0.0003
	(0.008)	(0.008)	(0.008)	(0.007)
Q2 cog. load	0.001	0.0004	0.002	-0.003
	(0.008)	(0.008)	(0.008)	(0.007)
Q1 cog. load	0.043^{***}	0.012	0.013	0.002
	(0.008)	(0.009)	(0.009)	(0.008)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.48	0.48	0.48	0.48
N	48,211	48,211	48,211	48,211
R^2	0.345	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on quintiles of the leave-out measure of cognitive load. The unit of observation is the encounter, considering 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%).

F.1.2 Logarithmic Cognitive Load

Table F.4: OLS Estimates of the Impact of Cognitive Load on Orders: Logarithmic Cognitive Load

	Number of diagnostic orders		Number medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1 + Cognitive load)	0.036***	0.020***	0.009***	0.005*	−0.002***	−0.002***
	(0.005)	(0.006)	(0.003)	(0.003)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.30	1.30	0.66	0.66	0.06	0.06
N	266,625	266,625	266,625	266,625	266,625	266,625
R^2	0.042	0.052	0.024	0.057	0.012	0.017
Adjusted R^2	0.039	0.047	0.021	0.052	0.009	0.012

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of orders that yield an abnormal result on the natural logarithm of the leave-out measure of cognitive load. The unit of observation is the *encounter–event* level, where an *action* either 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), (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%).

Table F.5: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking: Logarithmic Cognitive Load

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1 + Cognitive load)	0.059***	0.041***	−0.023***	−0.021***	−0.133***	−0.065***
	(0.004)	(0.005)	(0.002)	(0.002)	(0.005)	(0.005)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.045	0.054	0.022	0.032	0.020	0.112
Adjusted R^2	0.042	0.049	0.019	0.027	0.017	0.107

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 natural logarithm 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%).

Table F.6: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision: Logarithmic Cognitive Load

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
log(1 + Cognitive load)	-0.018^{***}	-0.006^{*}	-0.006^{*}	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.48	0.48	0.48	0.48
N	48,211	48,211	48,211	48,211
R^2	0.345	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the natural logarithm of the leave-out measure of cognitive load. The unit of observation is the encounter, considering 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%).

F.1.3 Logarithmic Number of Orders

Table F.7: OLS Estimates of the Impact of Cognitive Load on Logarithmic Orders

	log(1 + Number of diagnostic orders)		log(1 + Number of medication orders)	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.017***	0.011***	0.011***	0.005***
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.30	1.30	0.66	0.66
N	266,609	266,609	266,609	266,609
R^2	0.054	0.061	0.033	0.049
Adjusted R^2	0.051	0.056	0.030	0.044

Notes: The table presents OLS estimates of linear regressions of the logarithmic number of diagnostic and medication orders on the leave-out measure of cognitive load. The unit of observation is the *encounter–event* level, where an *action* either 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), (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%).

Table F.8: OLS Estimates of the Impact of Cognitive Load on Logarithmic Frequent Orders

	log(1 + Number of top25 DX orders)		log(1 + Number of non-top25 DX orders)	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.024***	0.019***	−0.001*	−0.002**
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.91	0.91	0.14	0.14
N	266,625	266,625	266,625	266,625
R^2	0.054	0.061	0.015	0.026
Adjusted R^2	0.052	0.056	0.013	0.021

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders and the number of orders outside the 25 most frequently placed 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%).

F.2 First-Order Batch

F.2.1 Orders Placed Within 20 minutes

Table F.9: OLS Estimates of the Impact of Cognitive Load on Orders Placed Within First 20 Minutes

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.200***	0.024	0.026***	−0.015**	0.006***	0.001
	(0.020)	(0.021)	(0.006)	(0.007)	(0.001)	(0.002)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	4.07	4.07	0.86	0.86	0.26	0.26
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.249	0.274	0.105	0.136	0.174	0.194
Adjusted R^2	0.241	0.260	0.096	0.119	0.166	0.179

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders placed within the first 20 minutes of the first physician-patient interaction and of the share of orders placed within this window that yield an abnormal result on the leave-out measure of cognitive load. The unit of observation is at the *encounter* level. 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%).

Table F.10: OLS Estimates of the Impact of Cognitive Load on Frequent Orders Placed Within First 20 Minutes

	Number of top25 DX orders		Number of non-top25 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.211^{***}	0.044^{**}	-0.011	-0.020^{**}
	(0.016)	(0.017)	(0.009)	(0.009)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	3.15	3.15	0.92	0.92
N	76,001	76,001	76,001	76,001
R^2	0.265	0.294	0.068	0.088
Adjusted R^2	0.258	0.281	0.058	0.071

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders and the number of orders outside the 25 most frequently placed orders, placed within the first 20 minutes of the first physician-patient interaction, on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is at the *encounter* level. 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%).

F.2.2 Orders Placed Within 30 minutes

Table F.11: OLS Estimates of the Impact of Cognitive Load on Orders Placed Within First 30 Minutes

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.217***	0.033	0.033***	−0.015**	0.005***	0.001
	(0.021)	(0.022)	(0.007)	(0.008)	(0.001)	(0.002)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	4.36	4.36	0.99	0.99	0.26	0.26
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.263	0.288	0.112	0.142	0.175	0.195
Adjusted R^2	0.255	0.275	0.103	0.126	0.166	0.180

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders placed within the first 30 minutes of the first physician-patient interaction and of the share of orders placed within this window that yield an abnormal result on the leave-out measure of cognitive load. The unit of observation is at the *encounter* level. 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%).

Table F.12: OLS Estimates of the Impact of Cognitive Load on Frequent Orders Placed Within First 30 Minutes

	Number of top25 DX orders		Number of non-top25 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.222***	0.050***	−0.006	−0.016*
	(0.017)	(0.018)	(0.009)	(0.010)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	3.35	3.35	1.01	1.01
N	76,001	76,001	76,001	76,001
R^2	0.280	0.310	0.074	0.095
Adjusted R^2	0.273	0.297	0.065	0.078

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders and the number of orders outside the 25 most frequently placed orders, placed within the first 30 minutes of the first physician-patient interaction, on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is at the *encounter* level. 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%).

F.2.3 Orders Placed Within 60 minutes

Table F.13: OLS Estimates of the Impact of Cognitive Load on Orders Placed Within First 60 Minutes

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.231***	0.038	0.045***	−0.014	0.005***	0.001
	(0.022)	(0.024)	(0.008)	(0.009)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	4.99	4.99	1.33	1.33	0.26	0.26
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.284	0.312	0.125	0.153	0.172	0.192
Adjusted R^2	0.276	0.300	0.115	0.138	0.163	0.177

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders placed within the first 60 minutes of the first physician-patient interaction and of the share of orders placed within this window that yield an abnormal result on the leave-out measure of cognitive load. The unit of observation is at the *encounter* level. 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%).

Table F.14: OLS Estimates of the Impact of Cognitive Load on Frequent Orders Placed Within First 60 Minutes

	Number of top25 DX orders		Number of non-top25 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.233***	0.054***	−0.002	−0.016
	(0.017)	(0.019)	(0.010)	(0.011)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	3.77	3.77	1.22	1.22
N	76,001	76,001	76,001	76,001
R^2	0.303	0.335	0.087	0.109
Adjusted R^2	0.296	0.323	0.077	0.092

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders and the number of orders outside the 25 most frequently placed orders, placed within the first 60 minutes of the first physician-patient interaction, on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is at the *encounter* level. 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%).

F.3 Alternative Cognitive Load Windows

F.3.1 60 Minutes Window

Table F.15: OLS Estimates of the Impact of Cognitive Load Based on 60-Minutes Window on Orders

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.) (60 min.)	0.052***	0.036***	0.004*	−0.002	0.001*	0.001**
	(0.006)	(0.006)	(0.002)	(0.003)	(0.001)	(0.001)
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
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.30	1.30	0.66	0.66	0.06	0.06
N	266,625	266,625	266,625	266,625	266,625	266,625
R^2	0.042	0.052	0.024	0.057	0.012	0.017
Adjusted R^2	0.040	0.047	0.021	0.052	0.009	0.012

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The relevant window for the computation of cognitive load is now 60 minutes (instead of 90 minutes, as in the standard specification). 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), (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%).

Table F.16: OLS Estimates of the Impact of Cognitive Load Based on 60-Minutes Window 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.) (60 min.)	0.073^{***}	0.056^{***}	-0.021^{***}	-0.021^{***}	-0.153^{***}	-0.067^{***}
	(0.005)	(0.005)	(0.002)	(0.002)	(0.004)	(0.004)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.046	0.054	0.022	0.032	0.023	0.112
Adjusted R^2	0.043	0.049	0.019	0.027	0.019	0.107

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 relevant window for the computation of cognitive load is now 60 minutes (instead of 90 minutes as in the standard specification). 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%).

Table F.17: OLS Estimates of the Impact of Cognitive Load Based on 60-Minutes Window on Diagnostic Precision

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.) (60 min.)	-0.012^{***}	-0.006^{*}	-0.006^{**}	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.48	0.48	0.48	0.48
N	48,211	48,211	48,211	48,211
R^2	0.344	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

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 relevant window for the computation of cognitive load is now 60 minutes (instead of 90 minutes as in the standard specification). 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%).

F.3.2 120 Minutes Window

Table F.18: OLS Estimates of the Impact of Cognitive Load Based on 120-Minutes Window on Orders

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.) (120 min.)	0.077***	0.052***	0.015***	0.006**	−0.0002	0.001
	(0.006)	(0.006)	(0.003)	(0.003)	(0.001)	(0.001)
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
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.30	1.30	0.66	0.66	0.06	0.06
N	266,625	266,625	266,625	266,625	266,625	266,625
R^2	0.043	0.052	0.024	0.057	0.012	0.017
Adjusted R^2	0.040	0.047	0.021	0.052	0.009	0.012

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The relevant window for the computation of cognitive load is now 120 minutes (instead of 90 minutes as in the standard specification). 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), (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%).

Table F.19: OLS Estimates of the Impact of Cognitive Load Based on 120-Minutes Window 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.) (120 min.)	0.096***	0.070***	−0.019***	−0.018***	−0.194***	−0.078***
	(0.005)	(0.006)	(0.002)	(0.003)	(0.005)	(0.005)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.046	0.054	0.022	0.032	0.023	0.112
Adjusted R^2	0.043	0.049	0.019	0.027	0.019	0.107

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 relevant window for the computation of cognitive load is now 120 minutes (instead of 90 minutes as in the standard specification). 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%).

Table F.20: OLS Estimates of the Impact of Cognitive Load Based on 120-Minutes Window on Diagnostic Precision

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.) (120 min.)	-0.019^{***}	-0.008^{**}	-0.008^{**}	-0.002
	(0.003)	(0.004)	(0.004)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.48	0.48	0.48	0.48
N	48,211	48,211	48,211	48,211
R^2	0.344	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

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 relevant window for the computation of cognitive load is now 120 minutes (instead of 90 minutes as in the standard specification). 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%).

F.4 Cognitive Load Based on Past Orders Only

Table F.21: OLS Estimates of the Impact of Cognitive Load (Based on Past Orders Only) on Orders

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.073***	0.048***	0.005**	0.003	−0.0003	0.0002
	(0.006)	(0.006)	(0.003)	(0.003)	(0.001)	(0.001)
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
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.30	1.30	0.66	0.66	0.06	0.06
N	266,625	266,625	266,625	266,625	266,625	266,625
R^2	0.042	0.052	0.024	0.057	0.012	0.017
Adjusted R^2	0.040	0.047	0.021	0.052	0.009	0.012

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). Only orders (instead of orders *and* notes as in the standard specification) within the last 90 minutes enter the measure cognitive load. 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), (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%).

Table F.22: OLS Estimates of the Impact of Cognitive Load (Based on Past Orders Only) 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.)	0.091***	0.065***	−0.018***	−0.017***	−0.138***	−0.047***
	(0.005)	(0.005)	(0.002)	(0.002)	(0.004)	(0.004)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.045	0.054	0.022	0.032	0.023	0.111
Adjusted R^2	0.043	0.049	0.019	0.027	0.020	0.107

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). Only orders (instead of orders *and* notes as in the standard specification) within the last 90 minutes enter the measure 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%).

Table F.23: OLS Estimates of the Impact of Cognitive Load (Based on Past Orders Only) on Diagnostic Precision

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.018^{***}	-0.007^{**}	-0.007^{**}	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.48	0.48	0.48	0.48
N	48,211	48,211	48,211	48,211
R^2	0.344	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

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). Only orders (instead of orders *and* notes as in the standard specification) within the last 90 minutes enter the measure cognitive load. 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%).

F.5 All Acting Physicians

Table F.24: OLS Estimates of the Impact of Cognitive Load on Orders: All Acting Physicians

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.068^{***}	-0.013^{***}	-0.013^{***}	-0.002	-0.009^{***}	-0.012^{***}
	(0.004)	(0.005)	(0.002)	(0.002)	(0.0004)	(0.0005)
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
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.19	1.19	0.66	0.66	0.07	0.07
N	472,166	472,166	472,166	472,166	472,166	472,166
R^2	0.030	0.052	0.016	0.108	0.015	0.029
Adjusted R^2	0.028	0.047	0.014	0.103	0.014	0.023

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result 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. The estimation now includes all actions *by any physician* (instead of just the primary physician as in the standard specifications). 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%).

Table F.25: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking: All Acting Physicians

	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.)	0.102***	0.029***	−0.034***	−0.042***	−0.147***	−0.062***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.004)	(0.004)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.81	0.81	0.38	0.38	429.33	429.33
N	472,166	472,166	472,166	472,166	272,735	272,735
R^2	0.035	0.058	0.015	0.032	0.027	0.108
Adjusted R^2	0.033	0.052	0.013	0.026	0.024	0.104

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 estimation now includes all actions *by any physician* (instead of just the primary physician as in the standard specifications). 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%).

Table F.26: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision: All Acting Physicians

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.018^{***}	-0.007^{**}	-0.007^{**}	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.48	0.48	0.48	0.48
N	48,211	48,211	48,211	48,211
R^2	0.344	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

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 estimation now includes all actions *by any physician* (instead of just the primary physician as in the standard specifications). 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%).

F.6 Varying Threshold for Common Orders

F.6.1 10 Most Frequent Orders

Table F.27: OLS Estimates of the Impact of Cognitive Load on Top10 Frequent Orders

	Number of top10 DX orders		Number of non-top10 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.056***	0.046***	0.015***	0.003
	(0.003)	(0.003)	(0.004)	(0.004)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.57	0.57	0.72	0.72
N	266,625	266,625	266,625	266,625
R^2	0.045	0.053	0.028	0.039
Adjusted R^2	0.042	0.048	0.025	0.034

Notes: The table presents OLS estimates of linear regressions of the number of the 10 most frequently placed orders and the number of orders outside the 10 most frequently placed 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%).

F.6.2 50 Most Frequent Orders

Table F.28: OLS Estimates of the Impact of Cognitive Load on Top50 Frequent Orders

	Number of top50 DX orders		Number of non-top50 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.083^{***}	0.063^{***}	-0.012^{***}	-0.013^{***}
	(0.005)	(0.006)	(0.002)	(0.002)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.05	1.05	0.24	0.24
N	266,625	266,625	266,625	266,625
R^2	0.045	0.054	0.027	0.036
Adjusted R^2	0.043	0.049	0.025	0.031

Notes: The table presents OLS estimates of linear regressions of the number of the 50 most frequently placed orders and the number of orders outside the 50 most frequently placed 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%).

F.6.3 100 Most Frequent Orders

Table F.29: OLS Estimates of the Impact of Cognitive Load on Top100 Frequent Orders

	Number of top100 DX orders		Number of non-top100 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.080***	0.060***	−0.010***	−0.010***
	(0.005)	(0.006)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.16	1.16	0.14	0.14
N	266,625	266,625	266,625	266,625
R^2	0.047	0.056	0.034	0.040
Adjusted R^2	0.044	0.051	0.031	0.035

Notes: The table presents OLS estimates of linear regressions of the number of the 100 most frequently placed orders and the number of orders outside the 100 most frequently placed 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%).

F.7 Individual Chief Complaints

F.7.1 Abdominal Pain

Table F.30: OLS Estimates of the Impact of Cognitive Load on Orders for Encounters with Abdominal Pain

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.066***	0.053***	0.027***	0.007	0.001	0.002
	(0.018)	(0.020)	(0.007)	(0.008)	(0.002)	(0.002)
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
Acuity code - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.52	1.52	0.72	0.72	0.06	0.06
N	33,185	33,185	33,185	33,185	33,185	33,185
R^2	0.035	0.056	0.007	0.036	0.005	0.023
Adjusted R^2	0.033	0.042	0.004	0.022	0.003	0.008

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the leave-out measure of cognitive load (standardized to mean 0 and SD 1), for the subset of encounters with chief complaint "Abdominal Pain." 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), (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%).

Table F.31: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking for Encounters with Abdominal Pain

	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.)	0.125***	0.099***	−0.059***	−0.046***	−0.200***	−0.090***
	(0.017)	(0.018)	(0.007)	(0.008)	(0.014)	(0.014)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.14	1.14	0.39	0.39	435.53	435.53
N	33,185	33,185	33,185	33,185	25,715	25,715
R^2	0.037	0.056	0.010	0.036	0.024	0.124
Adjusted R^2	0.034	0.043	0.008	0.022	0.021	0.109

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), for the subset of encounters with chief complaint "Abdominal Pain." 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%).

Table F.32: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision for Encounters with Abdominal Pain

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.021^{***}	-0.008	-0.009	-0.003
	(0.005)	(0.006)	(0.006)	(0.005)
Number of DX orders			-0.010^{***}	-0.009^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.267^{***}
				(0.067)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.24	0.24	0.24	0.24
N	6,837	6,837	6,837	6,837
R^2	0.025	0.137	0.151	0.387
Adjusted R^2	0.013	0.076	0.091	0.343

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), for the subset of encounters with chief complaint "Abdominal Pain". 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%).

F.7.2 Chest Pain

Table F.33: OLS Estimates of the Impact of Cognitive Load on Orders for Encounters with Chest Pain

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.105***	0.103***	−0.028**	−0.020*	−0.001	−0.001
	(0.022)	(0.024)	(0.011)	(0.011)	(0.002)	(0.002)
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
Acuity code - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.44	1.44	0.59	0.59	0.04	0.04
N	17,295	17,295	17,295	17,295	17,295	17,295
R^2	0.037	0.070	0.010	0.096	0.006	0.032
Adjusted R^2	0.032	0.045	0.005	0.071	0.001	0.006

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the leave-out measure of cognitive load (standardized to mean 0 and SD 1), for the subset of encounters with chief complaint "Chest Pain." 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), (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%).

Table F.34: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking for Encounters with Chest Pain

	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.)	0.126^{***}	0.116^{***}	-0.020^{**}	-0.013	-0.181^{***}	-0.080^{***}
	(0.020)	(0.022)	(0.008)	(0.008)	(0.017)	(0.017)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.11	1.11	0.32	0.32	464.20	464.20
N	17,295	17,295	17,295	17,295	15,542	15,542
R^2	0.041	0.070	0.006	0.046	0.026	0.137
Adjusted R^2	0.036	0.046	0.001	0.020	0.020	0.114

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), for the subset of encounters with chief complaint "Chest Pain." 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%).

Table F.35: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision for Encounters with Chest Pain

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.022^{**}	0.001	0.005	0.004
	(0.009)	(0.010)	(0.010)	(0.009)
Number of DX orders			-0.026^{***}	-0.027^{***}
			(0.002)	(0.002)
Share of top50 DX orders				-1.913^{***}
				(0.173)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.16	0.16	0.16	0.16
N	3,452	3,452	3,452	3,452
R^2	0.047	0.192	0.239	0.431
Adjusted R^2	0.023	0.082	0.134	0.352

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), for the subset of encounters with chief complaint "Chest Pain". 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%).

G Alternative Regression Specifications: Encounter Level

G.1 Max Cognitive Load

Table G.1: OLS Estimates of the Impact of Cognitive Load on Orders at the Encounter-Level

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Max cognitive load (std.)	0.701***	0.604***	0.345***	0.321***	0.007***	0.008***
	(0.022)	(0.024)	(0.012)	(0.013)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X	X	X
Max hours since shift start - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	4.55	4.55	2.32	2.32	0.09	0.09
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.201	0.236	0.142	0.211	0.046	0.066
Adjusted R^2	0.193	0.222	0.134	0.197	0.037	0.050

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.2: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking at the Encounter-Level

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Max cognitive load (std.)	0.531***	0.423***	0.171***	0.181***	0.264***	0.135***
	(0.017)	(0.019)	(0.010)	(0.011)	(0.011)	(0.010)
Prior cognitive load - Control	X	X	X	X	X	X
Max hours since shift start - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	3.20	3.20	1.35	1.35	1355.24	1355.24
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.201	0.236	0.102	0.134	0.061	0.361
Adjusted R^2	0.193	0.222	0.093	0.119	0.052	0.350

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 maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.3: OLS Estimates of the Impact of Cognitive Load on Entropy Reduction at the Encounter-Level

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Max cognitive load (std.)	-0.035^{***}	-0.026^{***}	-0.019^{***}	-0.017^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Number of diagnostic orders			-0.014^{***}	-0.007^{***}
			(0.0005)	(0.0005)
Share of top50 DX orders				-0.344^{***}
				(0.007)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.51	0.51	0.51	0.51
N	57,048	57,048	57,048	57,048
R^2	0.391	0.411	0.420	0.447
Adjusted R^2	0.384	0.398	0.407	0.434

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.4: OLS Estimates of the Impact of Deciles of Cognitive Load on Number of DX Orders and Inpatient Admission at the Encounter

	Number of diagnostic orders		Inpatient admission	
	(1)	(2)	(3)	(4)
Max cog. load decile 2	0.392***	0.198***	−0.006	0.008
	(0.062)	(0.070)	(0.005)	(0.006)
Max cog. load decile 3	0.841***	0.484***	0.001	0.007
	(0.065)	(0.073)	(0.006)	(0.007)
Max cog. load decile 4	1.116***	0.756***	0.012**	0.022***
	(0.067)	(0.075)	(0.006)	(0.007)
Max cog. load decile 5	1.363***	0.933***	0.014**	0.016**
	(0.070)	(0.077)	(0.006)	(0.007)
Max cog. load decile 6	1.566***	1.155***	0.026***	0.030***
	(0.072)	(0.078)	(0.006)	(0.007)
Max cog. load decile 7	1.762***	1.312***	0.038***	0.037***
	(0.074)	(0.080)	(0.007)	(0.007)
Max cog. load decile 8	1.961***	1.513***	0.044***	0.044***
	(0.076)	(0.082)	(0.007)	(0.007)
Max cog. load decile 9	2.056***	1.615***	0.053***	0.052***
	(0.079)	(0.085)	(0.007)	(0.008)
Max cog. load decile 10	2.341***	1.930***	0.067***	0.067***
	(0.085)	(0.092)	(0.008)	(0.008)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.55	4.55	0.24	0.24
N	76,001	76,001	76,001	76,001
R^2	0.202	0.236	0.221	0.242
Adjusted R^2	0.195	0.222	0.213	0.228

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic orders and a dummy variable whether the patient was admitted to the hospital during the encounter on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

G.2 Mean Cognitive Load

Table G.5: OLS Estimates of the Impact of Mean Cognitive Load on Orders at the Encounter-Level

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Mean cognitive load (std.)	0.285***	−0.005	0.122***	−0.023*	0.003**	0.001
	(0.020)	(0.023)	(0.011)	(0.012)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X	X	X
Max hours since shift start - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	4.55	4.55	2.32	2.32	0.09	0.09
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.197	0.243	0.141	0.219	0.047	0.067
Adjusted R^2	0.190	0.229	0.133	0.205	0.038	0.050

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.6: OLS Estimates of the Impact of Mean Cognitive Load on Frequent Orders and Note Taking at the Encounter-Level

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Mean cognitive load (std.)	0.275***	0.037**	0.011	−0.042***	0.166***	0.002
	(0.016)	(0.018)	(0.009)	(0.011)	(0.011)	(0.009)
Prior cognitive load - Control	X	X	X	X	X	X
Max hours since shift start - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	3.20	3.20	1.35	1.35	1355.24	1355.24
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.193	0.239	0.108	0.143	0.050	0.360
Adjusted R^2	0.185	0.225	0.100	0.127	0.041	0.349

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 mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.7: OLS Estimates of the Impact of Mean Cognitive Load on Entropy Reduction at the Encounter-Level

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Mean cognitive load (std.)	-0.026^{***}	-0.013^{***}	-0.013^{***}	-0.011^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Number of diagnostic orders			-0.014^{***}	-0.008^{***}
			(0.0005)	(0.0005)
Share of top50 DX orders				-0.345^{***}
				(0.007)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.51	0.51	0.51	0.51
N	57,048	57,048	57,048	57,048
R^2	0.389	0.410	0.420	0.446
Adjusted R^2	0.382	0.397	0.407	0.434

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.8: OLS Estimates of the Impact of Mean Cognitive Load on Inpatient Admission and Readmission at the Encounter-Level

	Inpatient admission		Readmission within 30 days	
	(1)	(2)	(3)	(4)
Mean cognitive load (std.)	0.001	-0.010^{***}	-0.002^{*}	-0.003^{***}
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.24	0.24	0.04	0.04
N	76,001	76,001	76,001	76,001
R^2	0.220	0.243	0.048	0.062
Adjusted R^2	0.212	0.229	0.039	0.045

Notes: The table presents OLS estimates of linear regressions of dummy variables whether the patient was admitted to the hospital during the encounter or re-admitted within 30 days on the mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.9: OLS Estimates of the Impact of Deciles of Mean Cognitive Load on Number of DX Orders and Inpatient Admission at the Encounter

	Number of diagnostic orders		Inpatient admission	
	(1)	(2)	(3)	(4)
Mean cog. load decile 2	1.358***	1.020***	0.024***	0.034***
	(0.069)	(0.070)	(0.006)	(0.006)
Mean cog. load decile 3	0.905***	0.357***	0.001	0.005
	(0.065)	(0.072)	(0.006)	(0.007)
Mean cog. load decile 4	1.657***	0.960***	0.025***	0.022***
	(0.069)	(0.074)	(0.006)	(0.007)
Mean cog. load decile 5	1.506***	0.729***	0.021***	0.016**
	(0.069)	(0.075)	(0.006)	(0.007)
Mean cog. load decile 6	1.518***	0.678***	0.014**	0.003
	(0.070)	(0.077)	(0.006)	(0.007)
Mean cog. load decile 7	1.637***	0.724***	0.032***	0.018**
	(0.071)	(0.078)	(0.006)	(0.007)
Mean cog. load decile 8	1.557***	0.580***	0.017**	−0.0005
	(0.072)	(0.080)	(0.007)	(0.007)
Mean cog. load decile 9	1.603***	0.579***	0.021***	−0.0004
	(0.074)	(0.083)	(0.007)	(0.008)
Mean cog. load decile 10	1.519***	0.440***	0.008	−0.017**
	(0.078)	(0.089)	(0.007)	(0.008)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.55	4.55	0.24	0.24
N	76,001	76,001	76,001	76,001
R^2	0.204	0.246	0.221	0.243
Adjusted R^2	0.196	0.232	0.213	0.229

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic orders and a dummy variable whether the patient was admitted to the hospital during the encounter on decile indicators of mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

G.3 All Acting Physicians

The following specifications include the actions by *all* physicians who interact with the patient encounter while attributing these actions to the primary note editor.

Table G.10: OLS Estimates of the Impact of Cognitive Load on Orders at the Encounter-Level: All Acting Physicians

	Number of diagnostic orders		Number of medication orders		Share of orders with abnormal test result	
	(1)	(2)	(3)	(4)	(5)	(6)
Max cognitive load (std.)	0.797***	0.679***	0.494***	0.522***	0.004***	0.004***
	(0.028)	(0.030)	(0.023)	(0.024)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X	X	X
Max hours since shift start - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	7.34	7.34	4.09	4.09	0.12	0.12
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.332	0.361	0.220	0.241	0.074	0.095
Adjusted R^2	0.325	0.349	0.212	0.228	0.065	0.079

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.11: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking at the Encounter-Level: All Acting Physicians

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Max cognitive load (std.)	0.588***	0.465***	0.209***	0.213***	0.039***	−0.006
	(0.020)	(0.020)	(0.015)	(0.016)	(0.009)	(0.007)
Prior cognitive load - Control	X	X	X	X	X	X
Max hours since shift start - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	4.99	4.99	2.36	2.36	1464.03	1464.03
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.348	0.382	0.165	0.185	0.025	0.287
Adjusted R^2	0.342	0.371	0.157	0.171	0.015	0.274

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 maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.12: OLS Estimates of the Impact of Cognitive Load on Entropy Reduction at the Encounter-Level: All Acting Physicians

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Max cognitive load (std.)	-0.041^{***}	-0.031^{***}	-0.027^{***}	-0.021^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Number of diagnostic orders			-0.007^{***}	-0.010^{***}
			(0.0003)	(0.0003)
Share of top50 DX orders				-1.103^{***}
				(0.010)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.51	0.51	0.51	0.51
N	57,048	57,048	57,048	57,048
R^2	0.393	0.412	0.417	0.529
Adjusted R^2	0.386	0.399	0.404	0.519

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.13: OLS Estimates of the Impact of Cognitive Load on Inpatient Admission and Readmission at the Encounter-Level: All Acting Physicians

	Inpatient admission		Readmission within 30 days	
	(1)	(2)	(3)	(4)
Max cognitive load (std.)	0.017***	0.013***	0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.24	0.24	0.04	0.04
N	76,001	76,001	76,001	76,001
R^2	0.244	0.264	0.051	0.065
Adjusted R^2	0.237	0.251	0.042	0.048

Notes: The table presents OLS estimates of linear regressions of dummy variables whether the patient was admitted to the hospital during the encounter or re-admitted within 30 days on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.14: OLS Estimates of the Impact of Deciles of Cognitive Load on Number of DX Orders and Inpatient Admission at the Encounter: All Acting Physicians

	Number of diagnostic orders		Inpatient admission	
	(1)	(2)	(3)	(4)
Max cog. load decile 2	0.843***	0.500***	−0.006	−0.0003
	(0.081)	(0.105)	(0.005)	(0.007)
Max cog. load decile 3	1.392***	0.915***	−0.011*	−0.004
	(0.086)	(0.108)	(0.006)	(0.007)
Max cog. load decile 4	1.758***	1.251***	−0.004	−0.002
	(0.091)	(0.111)	(0.006)	(0.007)
Max cog. load decile 5	2.048***	1.536***	−0.011*	−0.012*
	(0.095)	(0.112)	(0.006)	(0.007)
Max cog. load decile 6	2.369***	1.796***	0.004	−0.0005
	(0.097)	(0.114)	(0.006)	(0.007)
Max cog. load decile 7	2.455***	1.884***	0.016**	0.010
	(0.099)	(0.116)	(0.006)	(0.007)
Max cog. load decile 8	2.775***	2.204***	0.024***	0.015**
	(0.103)	(0.118)	(0.007)	(0.008)
Max cog. load decile 9	2.851***	2.308***	0.033***	0.024***
	(0.104)	(0.120)	(0.007)	(0.008)
Max cog. load decile 10	2.992***	2.471***	0.045***	0.034***
	(0.110)	(0.125)	(0.007)	(0.008)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	7.34	7.34	0.24	0.24
N	76,001	76,001	76,001	76,001
R^2	0.334	0.362	0.245	0.264
Adjusted R^2	0.327	0.350	0.237	0.251

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic orders and a dummy variable whether the patient was admitted to the hospital during the encounter on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).