Path Dependency in Physician Decisions

Lawrence Jin* Rui Tang† Han Ye‡ Junjian Yi§ Songfa Zhong¶

September 6, 2023

Abstract

We examine path dependency in physician decisions in an emergency department setting, and find that physicians' treatment decisions for the current and previous patients are positively correlated. We show that the positive autocorrelation is higher when the current patient is of greater medical uncertainty or more similar to the previous patient in terms of observed characteristics and when the physician is less experienced or more fatigued. We then show that these patterns are highly consistent with the memory and attention model (Bordalo, Gennaioli, and Shleifer, 2020), whereby the physician's current decision is anchored to her previous decision. The results from both reduced-form analyses and structural estimations provide further support for the importance of memory and attention in physician decision-making.

Keywords: physician decision; path dependency; memory; attention; anchoring and adjustment

JEL Codes: D91, I10

*Lee Kuan Yew School of Public Policy, National University of Singapore; email: lawrence-jin@gmail.com.

†Department of Economics, Hong Kong University of Science and Technology; email: ruitang.economics@gmail.com.

‡Department of Economics, Lingnan University; email: yehan0413@gmail.com.

§China Center for Economic Research, National School of Development, Peking University; email: junjian.yi@gmail.com.

¶Department of Economics, Hong Kong University of Science and Technology; Department of Economics, National University of Singapore; email: zhongsongfa@gmail.com.

"Of course, a doctor must know physiology and pathology and pharmacology. But he should also be schooled in heuristics—in the power and necessity of shortcuts, and in their pitfalls and dangers."

—Jerome Groopman, *How Doctors Think*

1 Introduction

The literature of behavioral economics has long suggested that individuals, including experts, use heuristic rules as shortcuts to make fast decisions, and yet these rules often lead to various cognitive biases (Kahneman, 2011). In the healthcare setting, due to medical emergencies and excess workloads, physicians often have to make fast decisions in a stressful environment with insufficient information. In this regard, as Groopman—a renowned physician at Harvard Medical School—states, it is important for physicians to understand the heuristic rules and the associated biases and errors. In the meantime, as Chandra, Handel, and Schwartzstein (2019) observe, "There has been relatively little research studying behavioral economics in the context of physician treatment decisions." Here we add to a small but growing literature on physician decisions from a behavioral perspective (Chan, 2016, 2018; Gong, 2017; Li, Dow, and Kariv, 2017; Chen, 2021; Silver, 2021).

We examine path dependency in physician decisions—that is, the effect of physicians' decisions with earlier patients on their decision for the current patient. Consider a physician who has just seen a patient with a common influenza. She may be more likely to diagnose influenza for the next patient with a cough, even if the patient actually has a rare lung disease. This example, provided by Chandra, Cutler, and Song (2011), illustrates that even though physicians should ideally make decisions based on the medical conditions of the current patient, they can be influenced by their previous treatment decisions.

The theoretical literature provides some guidelines for examining path dependency. In norm theory, Kahneman and Miller (1986) propose that similar past experiences triggered by the current event are consolidated into a norm that affects evaluation of the current event.[1] Based on a psychological mechanism of memory whereby more recent or more similar past experiences are more likely to be recalled (Kahana, 2012), Bordalo, Gennaioli, and Shleifer (2020) present a model of memory, attention, and choice. In their model, personal experiences or memories, especially ones that are more recent or more similar, serve as an anchor for decision makers to evaluate the current choice situation. This mechanism helps explain the previous example: The combination of recency and similarity in terms of coughing triggers the physician to be anchored to the previous patient who has a common influenza when treating the current patient, who actually has a rare lung disease. This can lead to positive autocorrelation in physician decisions.

Our primary empirical setting for examining path dependency in physician decisions is the emergency department (ED). This provides an ideal setting, because physicians make sequential and quick clinical judgments about patients with a wide variety of complaints and symptoms. They meet patients only once in most cases, and make their decisions based on limited information. We obtain administrative data for 253,466 patient visits to a large ED in a Southeast Asian country with treatments by 129 physicians over a period of 2 years (SAAH, 2013). The data contain comprehensive records on all ED visits, including patient characteristics, physician characteristics, physician decisions, and, importantly, timestamps for the patient's path through the ED. This dataset fits our research objective well, because the sequential order of patients assigned to the physician is conditionally random in this ED, which is critical for our identification.

Our primary finding is that ED physician decisions are on average positively autocorrelated. We focus on the physician's decision about whether to admit the patient for inpatient care. After controlling for patient characteristics, physician fixed effects, and time fixed effects, we find that the probability of admitting the current patient increases by 18.1% when the physician admitted the previous patient, compared with a situation in which she discharged the previous patient. Positive autocorrelation is also observed in

[1] When the event and the norm are similar, its evaluation is anchored to the norm; when the event and the norm are sufficiently different, its evaluation generates a surprise. Relatedly, in a prediction-error minimization framework (Hohwy, 2013), the neural system constructs models to predict the stimulus to receive, and the models are maintained as long as the discrepancy between the predicted and actual stimuli is not too large.

task-ordering decisions, ranging from ordering lab tests and imaging tests to prescribing opioids and antibiotics. Moreover, the estimated autocorrelation is higher when the current patient is of larger clinical uncertainty and when the physician is less experienced or more fatigued.

Our identification of path dependency hinges on two assumptions. The first assumption is that the order of patients assigned to the physician is random conditional on observable characteristics. This assumption is likely to be satisfied, since the sequential order of patients is determined by patients' arrival time and triage severity. We conduct nonparametric runs tests and find that patients are randomly ordered in observable characteristics. Moreover, the positive autocorrelation persists when we exclude consecutive patients who are of the same diagnostic category and who are from the same household or community. The second assumption is that the correlation in sequential decisions is not driven by common environmental factors shared by consecutive patients. We extensively investigate the second assumption. We demonstrate that (i) the results are robust after controlling for common environmental factors, including the presence of the same nurse or radiologist, colleagues' decisions, and hospital resource availability; (ii) the results are robust to the inclusion of the physician's and the ED's average admission rates in the current shift, which capture some unobserved environmental factors at physician and ED level; and (iii) the positive autocorrelation disappears when we randomly reassign patients among physicians within a given timeframe. These analyses provide consistent evidence that the positive autocorrelation in physician decisions cannot be attributed to nonrandom patient ordering or common environmental factors in the ED.

We test the generalizability of the observed path dependency in physician decisions in another important healthcare setting—cesarean section (C-section) decisions. Over the past two decades, the rate of C-sections has skyrocketed around the world (Card, Fenizia, and Silver, 2023). With almost 4 million babies born each year, childbirth is the most common cause for hospital admissions, and C-sections are the most common inpatient surgery in the US. We obtain administrative data for all 2,458,773 childbirths delivered by 3,105 physicians in the State of New York between 2005 and 2015.[2] In this setting, we also observe that C-section decisions are positively autocorrelated. Taken together, we portray a coherent and comprehensive picture of positive autocorrelation in physician decisions.

To explain our findings, we apply the theory of memory and attention (Bordalo, Gennaioli, and Shleifer, 2020) to our setting of path dependency in physician decisions. Specifically, when the physician sees the current patient, she retrieves similar past experiences, especially from the previous patient, to form a norm of clinical risk as an anchor. She then adjusts her perceived risk for the patient, depending on the difference between the norm and the current patient's risk. When the difference is small, the treatment decision for the current patient is anchored to that for the previous patient, leading to positive autocorrelation. Conversely, when the difference is unexpectedly large, the treatment decision for the current patient is adjusted excessively away from that for the previous patient, resulting in negative autocorrelation. The literature often refers to the first case of anchoring as assimilation and the second case of adjustment as the contrast effect.[3] Apart from path dependency, the anchoring and adjustment mechanism generates two additional predictions: The degree of autocorrelation between the current and earlier patients decreases with time distance and increases with similarity in characteristics.

Our reduced-form estimates are consistent with these predictions. Specifically, physician decisions are generally positively autocorrelated and the degree of autocorrelation is larger when the current and previous patients are closer in time or more similar in their characteristics. Notably, we also observe some evidence on negative autocorrelation when consecutive patients are similar in age and disease but sufficiently different in severity; in this case, the physician expects the two patients to have similar levels of risk, but is surprised by the difference in severity and thus overadjusts the perceived risk. Furthermore,

[2]This publication was produced from raw data purchased from or provided by the New York State Department of Health (NYSDOH, 2015). However, the conclusions derived, calculations, and views expressed herein are those of the authors and do not reflect the conclusions or views of NYSDOH. NYSDOH, its employees, officers, and agents make no representation, warranty or guarantee as to the accuracy, completeness, currency, or suitability of the information provided here.

[3]In our context of sequential decisions, we refer to the former as positive autocorrelation and the latter as negative autocorrelation.

we structurally estimate the model and find that the estimated model accommodates both positive and negative autocorrelation and is in line with the reduced-form evidence. Overall, both the reduced-form and structural findings provide empirical support for the memory and attention-based anchoring and adjustment mechanism (Bordalo, Gennaioli, and Shleifer, 2020).

Our study adds to the widely documented path dependency, which can be interpreted as supporting either assimilation or contrast effects in sequential decisions. Evidence of assimilation is documented in perceptual decisions (Akaishi et al., 2014); belief-updating tasks (Charness and Levin, 2005); judging the performance of Olympic gymnasts (Damisch, Mussweiler, and Plessner, 2006); lottery choice tasks (Erev and Haruvy, 2013); essay ratings (Zhao et al., 2017); and jury verdicts in criminal trials (Bindler and Hjalmarsson, 2019). The contrast effect is also supported in the literature, including movers from one city to another (Simonsohn and Loewenstein, 2006; Simonsohn, 2006); speed-dating participants (Bhargava and Fisman, 2014); interviewers (Radbruch and Schiprowski, 2022); and asylum judges, loan officers, and baseball umpires (Chen, Moskowitz, and Shue, 2016).

Overall, both positive and negative autocorrelation in sequential decisions can be accommodated by the framework of Bordalo, Gennaioli, and Shleifer (2020), whereby the sign of path dependency depends on decision contexts and the discrepancy between adjacent decision situations. We apply their framework and structurally estimate the model in the setting of physician decisions. Our empirical results demonstrate the pervasiveness of positive autocorrelation as well as the existence of negative autocorrelation in physician decisions, which suggests that the discrepancy between adjacent patients is generally within the range of the expectation of physicians, but it can occasionally be surprisingly large and lead to negative autocorrelation.

Our study also contributes to the burgeoning body of literature in health economics that focuses on physician decision-making. Physicians' performance is determined not only by their skills and productivity (Chandra and Staiger, 2007; Currie and MacLeod, 2017; Gong, 2017), but also by their preferences and surrounding environments (Chan, 2016; Li, Dow, and Kariv, 2017; Cutler et al., 2019). For instance, physicians are motivated by liability and financial incentives to perform unnecessary procedures (Currie, Gruber, and Fischer, 1995; Currie and MacLeod, 2008) and to report more higher-intensity service codes than actually delivered to receive higher Medicare reimbursements (Fang and Gong, 2017). In recent work, Chen (2021) finds that physicians' past collaboration raises team productivity. In an ED setting, Silver (2021) studies peer effects and finds that faster peers induce physicians to speed up and cut back on care. In another ED setting, Chan (2018) documents that physicians exhibit behavioral distortions near end of shift. Mullainathan and Obermeyer (2022) use machine learning to study physicians' decisions and find the coexistence of over- and undertesting, which indicates systematic errors in clinical judgment. Singh (2021) also finds sequential effects in the childbirth setting, whereby physicians are generally more likely to continue with the same delivery mode for the subsequent patient but switch to the other when the previous patient had complications with that delivery mode. Our study contributes to the literature by providing robust empirical evidence on path dependency in physician decisions and showing that our results can be explained by the behavioral mechanism of memory and attention (Bordalo, Gennaioli, and Shleifer, 2020).

2 Emergency Department and Administrative Data

2.1 Institutional Setting

Our primary empirical setting concerns physician decisions in the ED, which provides an interesting context to study sequential effects in physician decisions. First, ED physicians are general practitioners who treat patients with a wide variety of complaints and symptoms. Second, physicians make frequent clinical judgments and decisions, since patients are presented in quick succession. Third, ED physicians meet patients for the first time in most cases and their clinical decisions are based on limited information, such as the patient's chief complaints, symptoms, and demographic characteristics (Groopman, 2007). High patient loads and urgent patient needs require physicians to make sequential and quick clinical judgments based on limited information.

Our empirical analysis focuses on a large ED in a Southeast Asian country. This setting provides an ideal context for the study of sequential physician decisions, since the sequential order of patients assigned to the physician is conditionally random. Specifically, upon arrival at the ED, patients are assessed by a triage nurse and categorized into one of three severity levels: 1 refers to the most severe cases, 2 to major emergencies, and 3 to minor emergencies. Patients who have serious illness and injuries (severity levels 1 and 2; henceforth severe patients) are usually seen in the acute care area, and patients with mild conditions (severity level 3; henceforth non-severe patients) in the urgent care section.

The sequential order of treatment is determined by patients' triage severity level and arrival time only. A computer-based patient-scheduling system automatically assigns patients to available physicians. In the acute care area, severity level 1 patients have a higher priority than level 2 patients. Within each of the three severity levels, patients are sorted by time of arrival. Physicians generally attend to a new patient from the top of the patient queue, which ensures that for a given severity level, the earliest arrival is treated first. We validate this assignment policy using timestamps for patients' paths through the ED; in our data, almost all patients are sequentially treated based on their arrival time conditional on triage severity.[4] This institutional feature ensures that patient ordering is conditionally random, which is a substantial advantage because it validates our first identifying assumption which will be discussed in Section 3.2.

The studied ED further provides a clean context for our study. First, physicians' shifts are scheduled at least weeks in advance and are not publicly available; therefore, patients have no information regarding which physician will be on shift before they arrive at the ED. Second, physicians cannot control the volume of ED arrivals or the types of patients assigned to them during a shift. Third, physicians' decisions in the ED are not influenced by financial incentives. Physicians are paid a basic monthly salary with a fixed shift allowance, and are not directly rewarded for the quality or quantity of work during the scheduled shift. Finally, government subsidies are provided for every ED patient. All

[4]Results are robust when we exclude patient visits that are not sequentially treated based on their arrival time and triage severity.

patients incur a fixed attendance fee of around $86 US Dollars per visit, which covers the cost of consultation, medicine, procedures, and tests.

In summary, the ED provides an ideal setting in which the sequential order of patients assigned to each physician is conditionally random; the sequential order of treatment is determined entirely by patients' triage severity and arrival time. Because of the unexpected nature of ED visits, the order in which patients arrive at the ED is largely random. The internal shift schedule of physicians is predetermined. Neither patients nor physicians can select each other. Patients are thus assigned to the physician in a random order conditional on their triage severity level.

2.2 Administrative Data

We obtain administrative data for all patient visits to the ED from January 1, 2011, to December 31, 2012. The hospital information system documents comprehensive records for each visit, including patient characteristics, physician identifier, clinical decisions, and, importantly, timestamps for the patient's path through the ED. These records allow us to track real-time patient flow and the universe of physicians' activities in the ED (see Appendix A.1 for details).

Over the 2 years, we observe 264,115 patient visits to the ED. We limit our attention to physicians who have treated a minimum of 100 patients during the 2-year period. We also exclude visits in which the patient died upon arrival or in the ED, left before being seen, or self-discharged against medical advice. Finally, for the purpose of analysis, we restrict our sample to patient visits in which the immediately prior patient treated by the same physician occurred within 48 hours. These exclusions restrict the sample to 253,466 patient visits treated by 129 physicians.

Physician Decisions. Physicians in the ED provide immediate evaluation, care, and stabilization to patients; at the same time, they act as gatekeepers to inpatient specialist units. The disposition decision to admit or discharge a patient is the primary product of ED care and a matter of discretion for physicians (Chan, 2018). We use disposition decisions as our primary measure of physician decisions in the ED. We also investigate physician decisions on lab tests (testing for a sample of blood, urine, tissue, or other bodily substances); imaging tests (X-ray, ultrasound, computed tomography scan, and magnetic resonance imaging); opioid prescriptions; and antibiotic prescriptions. Panel A of Table 1 summarizes physician decisions in the ED. The average admission rate for inpatient care is 22%. Overall, 52% of patients undergo one or more lab tests, 51% receive imaging tests, 8% receive opioids, and 5% receive antibiotics.

Patient Characteristics. Our dataset records information available to the physician at the time of accepting the patient, including the patient's gender, age, race, and triage severity level. Panel B of Table 1 reports summary statistics for patient characteristics. In our data, 65% of patients who visited the ED during the sample period were men. Patients' average age was around 39 years.[5] Patient race is grouped into four broad categories, comprising 55% as Race 1, 20% as Race 2, 16% as Race 3, and the remaining 9% as other races. Around 71% of patients were severity level 3, 25% were level 2, and 4% were level 1. Because the data do not include a patient's chief complaint or reason for the ED visit, we capture patients' medical conditions using their diagnostic information. We codify the diagnostic information into 18 broad categories based on the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM).[6] Instead of using more granular categories, we believe the broad classification of 18 categories is less subject to physician discretion. Our results are robust when diagnostic categories are not controlled for and when more detailed categories are controlled for.

Patient arrivals at the ED exhibit considerable fluctuations across time. In particular, Sundays and Mondays were the busiest days within a week, and 10 am to 3 pm and 8 pm to 11 pm were the two peak hours within a day. The total number of patient visits increased over the 2 years, and ED patient volumes varied across months. We include a

[5]In our regression analyses in the ED setting, patients are divided into nine groups by age: 0–14, 15–19, 20–24, 25–34, 35–44, 45–54, 55–64, 65–74, and 75 and over.

[6]Diseases are classified into 18 broad categories: infectious and parasitic diseases; neoplasms; endocrine, nutritional, and metabolic diseases and immunity disorders; diseases of the blood and blood-forming organs; mental illness; diseases of the nervous system and sense organs; diseases of the circulatory system; diseases of the respiratory system; diseases of the digestive system; diseases of the genitourinary system; complications of pregnancy, childbirth, and the puerperium; diseases of the skin and subcutaneous tissue; diseases of the musculoskeletal system and connective tissue; congenital anomalies; certain conditions originating in the perinatal period; injury and poisoning; symptoms, signs, and ill-defined conditions and factors influencing health status; and residual codes, unclassified, and all e codes.

set of time fixed effects—hour of day, day of week, and month by year—in our regression analysis to account for time variations in patient visits.

3 Econometric Model and Identifying Assumptions

3.1 Econometric Model

We use the following baseline specification to estimate the causal effect of the physician's lagged decision on her decision for the current patient, conditional on characteristics of the current patient visit and physician fixed effects:

$$Y_{it} = \alpha_0 + \alpha_1 Y_{i,t-1} + X_{it}\beta + \delta_i + \mu_{it}, \tag{1}$$

where subscript i denotes a physician and t denotes the sequence of patient visits assigned to the physician. The dependent variable Y_{it} represents binary decisions (e.g., admit or not) by physician i for current patient t. The independent variable of interest $Y_{i,t-1}$ is the lagged dependent variable—that is, the physician's decision for the previous patient.

We control for characteristics of current patient visit X_{it}, including patient risk adjusters—patient demographics and medical conditions—and time fixed effects. Patient risk adjusters allow for the sorting of patient types across physicians. The inclusion of time fixed effects captures variations in patient flows, medical resource capacity, and unobserved time-specific patient characteristics. We also include physician fixed effects δ_i to control for physician heterogeneities. That is, the estimation in Equation (1) exploits within-physician variations instead of cross-physician variations in medical decisions. Finally, μ_{it} is the error term. Standard errors are clustered at physician level throughout the analyses.

Our coefficient of interest α_1 measures the change in the probability that the physician makes an affirmative decision for the current patient if her decision for the previous patient was affirmative rather than negative. Given the two identifying assumptions discussed below, the coefficient α_1 measures the causal effect of the lagged decision on the current decision. In particular, $\alpha_1 > 0$ ($\alpha_1 < 0$) indicates positive (negative) autocorrelation in physician decisions.

The literature has not reached a consensus on whether physician fixed effects should be included in estimating decision autocorrelation.[7] On the one hand, using panel data with heterogeneities across physicians, the estimate of α_1 is biased upward if we do not include physician fixed effects. For example, in the ED, the unobserved tendency to admit a patient is a physician-specific characteristic, which affects the chance of inpatient admission for all patients treated by the physician. If physician fixed effects are left in the error term, the estimate of α_1 would be upward biased because both the current and previous decisions are positively correlated with the physician's unobserved tendency for inpatient admission. In this case, if α_1 is positive, the pooled OLS estimate without controlling for physician fixed effects provides an upper bound for α_1. On the other hand, controlling for physician fixed effects in Equation (1) leads to a downward bias for the estimate of α_1, if α_1 is positive and the number of patients treated by each physician is small (Nickell, 1981).[8] This downward bias is termed the "Nickell bias" in the literature. With the Nickell bias, the fixed-effect estimate provides a lower bound for α_1.

We include physician fixed effects in estimating Equation (1) for three reasons. First, our estimates of α_1 are generally positive, as reported below. Even if the Nickell bias exists, our fixed-effect estimate based on Equation (1) provides a lower bound for the positive α_1. Second, the Nickell bias is minimized in our study, since we restrict our estimation sample to physicians who have treated at least 100 patients.[9] Finally, our results remain robust when we follow the literature and use alternative estimators to address the Nickell

[7]For example, Bhargava and Fisman (2014) and Bindler and Hjalmarsson (2019) include decision-maker fixed effects in their main specifications, whereas Chen, Moskowitz, and Shue (2016) use alternative controls for decision-maker heterogeneities instead of fixed effects.

[8]Consider the regression equation $y_{it} = \alpha_0 + \alpha_1 y_{i,t-1} + X_{it}\gamma + \delta_i + \mu_{it}$ $(i = 1, \ldots, N;\ t = 1, \ldots, T)$. The fixed-effect estimator is equivalent to the pooled OLS estimator for $\ddot{y}_{it} = \alpha_1\, \ddot{y}_{i,t-1} + \ddot{X}_{it}\gamma + \ddot{\mu}_{it}$, where $\ddot{z}_{it}(= z_{it} - \bar{z}_i)$ is the deviation of z_{it} $(\in \{y_{it},\ X_{it},\ \mu_{it}\})$ from $\bar{z}_i(= T^{-1}\sum_{r=1}^{T} z_{ir})$, and $\ddot{y}_{i,t-1} = y_{i,t-1} - T^{-1}\sum_{r=0}^{T-1} y_{ir}$. Nickell (1981) shows that when T is small, the pooled OLS estimator based on the demeaned equation is biased even if N, the number of individuals, goes to infinity. The bias arises because the correlation between $\ddot{y}_{i,t-1}$ and $\ddot{\mu}_{it}$ is nonzero. If $\alpha_1 > 0$, the bias is negative; the degree of positive autocorrelation is underestimated.

[9]Nickell (1981) demonstrates that the bias with the pooled OLS estimate of α_1 based on the demeaned equation is sizable when T is small, but the bias approaches zero when T is sufficiently large.

bias.[10]

3.2 Two Identifying Assumptions

The causal interpretation of α_1 in Equation (1) hinges on the assumption of no systematic sorting on unobserved factors that predict physician decisions. That is, the error term μ_{it} is not autocorrelated, and thus the variable of interest $Y_{i,t-1}$ is not correlated with μ_{it}. The error term captures both unobserved patient characteristics and environmental factors that may affect physician decisions. Hence, the causal interpretation of α_1 depends on two specific assumptions.

The first is the conditionally random ordering of patients. That is, the order of patients assigned to the physician is not systematically sorted on any unobserved characteristics that correlate with the probability of an affirmative decision. Imagine that some components of medical complications are unobserved and not perfectly correlated with our measures of patient characteristics. When patients' ordering is positively (negatively) correlated with such unobserved components, we would estimate a spurious positive (negative) correlation between the physician's current and previous decisions. If this were the case, the estimate of α_1 in Equation (1) would be confounded by the correlation in unobserved patient characteristics between the current and previous patients.

The second assumption is that the correlation in sequential medical decisions is not driven by common environmental factors shared by consecutive patient visits. Medical decisions are not solely determined by the physician in charge but also affected by surrounding environments, such as medical equipment availability and other medical staff—nurses, physician assistants, radiologists, etc.—working at the same time. The availability of the same medical resource capacity and the presence of the same staff are more likely between closer points in time than more distant points in time. Common environmental

[10] The literature has proposed two alternative estimators to address the Nickell bias when T is small. One solution is to take first differences $\Delta \mathrm{y}_{it} = \alpha_1 \Delta \mathrm{y}_{i,t-1} + \Delta X_{it}\gamma + \Delta\mu_{it}$. The first difference removes the individual fixed effects, but $\Delta\mu_{it}$ correlates with $\Delta \mathrm{y}_{i,t-1}$. The Anderson–Hsiao estimator uses $y_{i,t-2}$ as an instrument for $\Delta \mathrm{y}_{i,t-1}$ (Anderson and Hsiao, 1982). Another solution is to estimate $y_{it} = \alpha_0 + \alpha_1 y_{i,t-1} + X_{it}\gamma + c_i + \mu_{it}$ with an alternative control for individual heterogeneity c_i instead of fixed effects. For example, Chen, Moskowitz, and Shue (2016) control for individual heterogeneity using a recent moving average of the previous n decisions made by each decision maker. In Panels A and B of Appendix Table A.2, we show that our results remain robust using these two estimators.

factors may induce spurious correlation in physician decisions for patients treated close in time and confound the effect of the physician's previous decision on her current decision.

The literature on sequential decisions has largely investigated the first assumption of conditionally random ordering (e.g., Bhargava and Fisman (2014); Chen, Moskowitz, and Shue (2016); Bindler and Hjalmarsson (2019)). However, the second assumption regarding common environmental factors is less studied. In our empirical analysis, we will extensively examine both assumptions.

3.3 Empirical Specification Details and the Conditionally Random Ordering of Patient Visits

We test path dependency in physician decisions in the ED by estimating Equation (1). In our main regression analyses, we focus on physicians' disposition decisions. We define Y_{it} ($Y_{i,t-1}$) as an indicator for whether the physician admits the current (previous) patient for inpatient care. Control variables include, unless otherwise stated, patient demographics (gender, age, and race), triage severity, diagnoses, physician fixed effects, and time fixed effects. We estimate Equation (1) using a linear regression model, allowing for clustered standard errors. Our results remain robust when we use probit and logit models (Panels C and D, Appendix Table A.2).

We control for time fixed effects of hour of day, day of week, and month by year in our main regression. Controlling for time fixed effects is important. First, as discussed above, the volume of ED arrivals fluctuates over time. Second, hospital resource availability is highly time dependent. For example, one of the most critical resources—inpatient beds—has a higher occupancy rate in the morning than in the afternoon (Shi et al., 2015). Third, patient conditions may also vary across time. Some diseases have regular daily or seasonal variations in their risk level or severity of symptoms. For example, the risk of heart attacks is highest in the morning and asthma is worse at night than during the day (Litinski, Scheer, and Shea, 2009). Patients with certain types of diseases may consecutively visit the ED at specific times of the day. Also, ED visits at night or on the weekend could differ from those in the daytime or on weekdays, because regular outpatient care is unavailable at those times. The inclusion of time fixed effects captures variations in patient arrivals, medical resource capacity, and unobserved time-specific patient characteristics.

We conduct the nonparametric runs test on the randomness of observed patient characteristics. In a series of consecutive observations, a run is defined as a sequence in which a certain type of observation is repeated one or more times. A new run occurs each time the alternate type is observed. Consider a hypothetical physician who sees 10 patients: five males (M) and five females (F). If all male patients are presented first (MMMMMFFFFF), we say that there are two runs in the data—the minimum number of runs possible. If male and female patients are placed alternately (MFMFMFMFMF), there are 10 runs—the maximum number of runs possible. Obviously, both hypothetical sequences are not likely to be random on patient gender; there are either too few or too many runs in a sequence. The runs test detects whether a sequence of data occurs in a random process based on the number of runs. A small number of runs suggests positive serial correlation and a large number negative serial correlation.

For each physician-shift,[11] we separately test whether patient sequence follows a random order with respect to patient gender, race, age, and admission probability predicted by patient demographics, triage severity, and diagnostic categories. For each characteristic considered, the test produces a p-value for each physician-shift. The smaller the p-value, the stronger the evidence that we can reject the null hypothesis of randomness. Figure 1 presents the distribution of p-values from the runs test for each physician-shift. Across all panels, the fraction of shifts with low p-values (less than 0.05) is around 5%.[12] For

[11] Following the procedure of Brachet, David, and Drechsler (2012), we construct physicians' shifts based on their periods of inactivity, which is identified by their absence from the administrative data. Sorting data first by physician ID, then by the date and time during which physicians were involved in each patient visit, we define the beginning of a new shift when 6 or more hours have elapsed between consecutive observations of the same physician. The results we present below remain robust if we use 4-hour and 5-hour cutoffs to define new shifts.

[12] We note that the proportion of low p-values is slightly larger than 5% when testing randomness with respect to admission probability. This is because the order of patients is presumably nonrandom with respect to triage severity in shifts in which physicians cross-cover the two treatment areas. For example, a physician who is scheduled to work in the urgent care area may switch to the acute care area halfway through her shift, when there is an imbalance of patient visits. This yields a patient sequence in which all non-severe (low admission risk) patients are treated before severe (high admission risk) patients. In Appendix Figure A.3, we exclude shifts in which physicians cross-cover the two treatment areas and reexamine the randomness with respect to admission probability; only 4% of those shifts are associated with low p-values.

the vast majority of shifts, we do not find any evidence of nonrandom ordering based on patient gender, race, age, or admission probability. The results we present below are not sensitive when we exclude shifts with low p-values from the runs tests.

Our first identifying assumption of conditionally random ordering is very likely to be satisfied in the ED setting. The nonparametric runs tests provide consistent evidence that patients are randomly ordered in observed characteristics. In the robustness analysis below, we present evidence that our estimated autocorrelation in physician decisions is not attributable to autocorrelation in unobserved patient conditions or common environmental factors.

4 Results

We observe positive autocorrelation in a range of ED physician decisions by tabulating the raw data (Appendix Figure A.2). More specifically, the physician is more likely to admit the current patient if she admitted the previous patient compared with if discharged the previous patient. Moreover, the physician is more likely to order lab tests (imaging tests) if she ordered lab tests (imaging tests) for the previous patient. Finally, the physician is also more likely to prescribe opioids (antibiotics) if she did so for the previous patient. In what follows, we examine path dependency using regression analyses, conduct extensive robustness checks, analyze heterogeneity across different subgroups, and explore the consequences of path dependency.

4.1 Main Results on Autocorrelation

We present the fixed-effect estimates from Equation (1) in Panel A of Table 2. Column (1) reports results for the full sample of patient dispositions, in which consecutive decisions are made by the same physician within 48 hours. The estimate of α_1 is positive and statistically significant. The probability of admitting the current patient is 3.2 percentage points higher if the physician admitted the previous patient, compared with if she discharged the previous patient. The estimate represents a 14.7% increase relative to the mean inpatient admission rate of 0.217. This indicates a strong positive autocorrelation in admission decisions conditional on patient demographics, triage severity, diagnoses, physician fixed effects, and time fixed effects.

We then divide the full sample into two subsamples. The first subsample are those who follow another patient within the same shift by excluding the first patient in each shift. In column (2), we show that the estimate is 0.038, which is statistically significant at the 1% level. The estimate represents a 18.1% increase in admission decisions for patients who follow another patient in the same shift. The second subsample are those who are the first patient in each shift. In contrast to the estimate in column (2), the estimate in column (3) is small in magnitude and statistically insignificant.[13] The physician's disposition decision for the last patient in the previous shift has no significant effect on her decision for the first patient in the current shift. The estimated positive autocorrelation in column (1) is driven by consecutive decisions that occur within the same shift.

Panel B of Table 2 reports pooled OLS estimates without controlling for physician fixed effects. Consistent with the literature (Nickell, 1981; Bindler and Hjalmarsson, 2019), the pooled OLS estimates are larger than the fixed-effect estimates. Using the full sample, column (1) in Panel B shows that a lagged admission decision increases the probability of a subsequent admission decision by 3.8 percentage points (17.6%). Column (2) shows that the effect increases to 4.4 percentage points (21.2%) if the current and previous patient are treated within the same shift. Again, decision autocorrelation disappears if the two consecutive patients were treated in different shifts (column (3)). Taking the fixed-effect estimate as the lower bound and the pooled OLS estimate as the upper bound, a lagged admission decision increases the probability of a subsequent admission in the same shift by 18.1% to 21.2%. For simplicity, in the remaining analyses in the ED setting, we focus on fixed-effect estimates based on observations in which the current and previous patient are treated within the same shift.

[13] We note that the mean admission rate in column (3) is higher than that in column (2). This is because severe patients have longer consultations than non-severe patients; a physician can treat more patients when her shift is in the urgent care section with non-severe patients, compared with when her shift is in the acute care area with severe patients. Thus, the share of severe patients of the first patients in a shift is larger than the share of severe patients of total patients.

Other Physician Decisions. We examine path dependency in other physician decisions by estimating Equation (1) with different measures of physician decisions. Column (1) of Table 3 shows that the probability that the physician orders any lab tests for the current patient is 2.3 percentage points higher when she ordered lab tests for the previous patient, compared with when she did not. Column (2) indicates that the probability that the physician orders any imaging tests is 3.0 percentage points higher if she ordered imaging tests on the previous patient. Column (3) shows that the probability that the physician prescribes opioids increases by 1.8 percentage points if she prescribed opioids for the previous patient. Finally, column (4) shows that the physician is 2.0 percentage points more likely to prescribe antibiotics for the current patient if she prescribed antibiotics for the previous patient. All four estimates are statistically significant at the 1% level, and support positive autocorrelation in physician decisions.

4.2 Robustness Analyses

Autocorrelation in Patient Conditions. As discussed in Section 2.1, the ED provides an ideal setting in which the sequential order of patients assigned to each physician is conditionally random. Patients are sequentially treated based on their arrival time, conditional on triage severity. Nonparametric runs tests provide evidence that patients are randomly ordered in observed characteristics. Here, we further rule out possibilities of nonrandom patient ordering. We take column (2) in Panel A of Table 2 as the baseline specification and, to ease comparison, report this again in column (1) of Appendix Table A.3. We show that our results are robust when we exclude consecutive patients who are likely to have autocorrelated conditions and after we control for proxies for unobserved medical conditions.

First, we consider circumstances in which patient arrivals at the ED may not be random. For instance, major accidents or public health events may bring in a large volume of patients with similar conditions at the same time, resulting in autocorrelated conditions in consecutive patients. This could generate spurious positive autocorrelation in sequential decisions.

We address the problem of autocorrelated patient conditions by estimating Equation (1) with restricted samples of patients. First, patients brought in from the same accident or public health event are likely to have the same (broad) diagnosis. We thus exclude patients who shared the same diagnostic category with the previous one. Second, patients from the same household or the same community may share some similar conditions. For example, infectious diseases are more likely to spread within family or community members. We exclude consecutive patients who are from the same household or from the same community. Columns (2)–(4) in Appendix Table A.3 show that estimates from the restricted samples remain positive and statistically significant. The results suggest that the positive autocorrelation in physician decisions is not driven by autocorrelation in patient conditions.

Second, to explore the potential role of unobserved variation in patient characteristics, we test the sensitivity of our results to controlling for richer measures of patient characteristics. We classify patient diagnoses in 285 categories according to Clinical Classifications Software, and control for the detailed 285 diagnostic categories instead of the 18 broad categories in the main regression. Although the detailed diagnosis is an imperfect (yet the best available) proxy measure of unobserved factors and is subject to physician decisions, it reflects more detailed information on patient conditions. The last two columns in Appendix Table A.3 show that our results remain largely unchanged regardless of whether we control for diagnostic categories and whether we control for 18 broad categories or 285 detailed categories. The results suggest that patients are not likely to sort sequentially on unobserved patient conditions.

Common Environmental Factors. We assess the second identifying assumption in this subsection. We find that the estimated positive autocorrelation in physician decisions remains robust after we address issues relating to various common environmental factors. We first consider the environmental factors observed in the data and demonstrate that (i) positive autocorrelation in physician decisions cannot be attributed to the presence of the same nurse or radiologist; (ii) positive autocorrelation also cannot be attributed to the decisions of other physicians who are currently treating other patients in the ED; and (iii) positive autocorrelation is robust to controlling for hospital resource availability. We also assess environmental factors that are unobserved in the data. We show that the results are robust to inclusion of the average admission rate for the physician in the current shift (excluding the current patient) and the average ED admission rate among all physicians in the current shift (excluding the physician who treats the current patient); these two variables are used to measure unobserved environmental factors shared with the current patient at physician level and ED level. We also show that the positive autocorrelation in physician decisions disappears when we randomly reassign patients among physicians within a given timeframe.

First, medical decisions are not only determined by physicians, but affected by other medical staff, such as nurses and radiologists, who also play an important role in patient care. The presence of the same medical staff is more likely between closer points in time than more distant points in time; this may contribute to similar medical decisions for patients treated close in time. To address this concern, we additionally control for fixed effects for both the main nurse and the main radiologist working with the physician in each shift.[14] In a second specification, we control for physician-main nurse group fixed effects or physician-main radiologist group fixed effects. Appendix Table A.4 requires that both the main nurse and the main radiologist are involved in at least 50 patients each, and column (1) repeats the baseline analysis without controlling for nurse or radiologist fixed effects using the restricted sample. Columns (2)–(4) show that the result remains robust in both specifications. The positive autocorrelation in physician decisions cannot be attributed to the presence of the same nurse or radiologist.

Second, physician decisions may be affected by the decisions of other physicians who are currently treating other patients in the ED. We further include the latest disposition decision made by other physicians in the ED as an independent variable, in addition to the main nurse and main radiologist fixed effects. Column (5) of Appendix Table A.4

[14]We infer nurse and radiologist information from task records. The hospital information system records the one who finishes each task (either the physician or the nurse for non-radiology tasks and the radiologist for radiology tasks). Nurse information is not available if the patient receives few or no tasks, and radiologist information is not available for patients without radiology tasks. We identify the main nurse (radiologist) who is most involved in each physician-shift. Our results are robust if we define the main nurse (radiologist) working with the physician in 2-hour or 3-hour intervals.

shows that the estimated positive autocorrelation remains unchanged after controlling for the latest disposition decision by colleagues. In addition, the coefficient on colleagues' disposition decision is small in magnitude and statistically insignificant. This suggests that physicians' current decisions are affected by their own decisions for previous patients, but not by the recent decisions of their colleagues. This is consistent with the findings of Silver (2021), who demonstrates that peer-induced pressure plays a limited role in admission decisions at ED level.

Third, physician decisions are constrained by medical resource capacity. For example, the decision to admit a patient is subject to the availability of inpatient beds. Physicians may reduce inpatient admissions under a high level of inpatient bed occupancy and admit more patients if more beds are available. If this were the case, we might observe similar disposition decisions in consecutive patients. As discussed in Section 3.3, the concern regarding bed availability can be absorbed partly by time fixed effects in our regression. To further address this concern, we estimate Equation (1) by including inpatient bed availability as an additional independent variable, in addition to the main nurse and main radiologist fixed effects and colleagues' latest decision. We use two variables to measure the level of bed occupancy: (i) total number of inpatient admissions issued in the ED in the last 12 hours[15] and (ii) average waiting time from the admission decision to an inpatient ward for patients whose consultation ended in the previous hour. Columns (6) and (7) of Appendix Table A.4 report regression results with each of the two measures. In both analyses, the estimated coefficients on bed occupancy are negative and small in magnitude and the estimated coefficient on the lagged admission remains largely unchanged. These results suggest that bed occupancy plays a limited role in physicians' admission decisions in the studied ED.

Another factor of interest is ED crowdedness, because physicians may adjust their behavior accordingly. For example, they may ration access to inpatient care and discharge more patients when demand surges. If this were the case, patients treated in the same

[15]The ED is the major source of inpatient admissions and accounts for around 70% of patients in the hospital's inpatient units. Results remain unchanged if we measure bed occupancy using the total number of inpatient admissions issued in the ED in the last 24 hours or 48 hours.

time period might receive similar dispositions. To address this concern, we additionally include controls for ED crowdedness as a robustness analysis. We measure the degree of crowdedness using a physician-adjusted value of system load.[16] This is defined as the ED system load divided by the number of physicians on staff during the current patient's consultation. The ED system load measures the total number of patients in the ED, including those waiting to be seen and those being treated. Column (8) of Appendix Table A.4 presents the estimation result after controlling for the adjusted system load. The coefficient on the lagged admission remains largely unchanged. This implies that ED crowdedness cannot explain the positive autocorrelation in physician decisions.

Next, we consider unobserved environmental factors shared by consecutive patient visits. We further control for the average admission rate for the physician in the current shift, calculated excluding the current patient. This variable captures potential trends in admissions, patient conditions, and other common environmental factors at physician-shift level. We also control for the average ED admission rate among all physicians during the current shift, calculated by excluding the physician who treats the current patient. This variable accounts for all observed and unobserved factors at ED level. Column (9) of Appendix Table A.4 reports the estimation results. The coefficients remain largely unchanged after controlling for the two variables.

Finally, we conduct placebo tests. We randomly reassign patients among physicians in each 4-hour interval[17] and estimate our main regression with the simulated data. We repeat this procedure 100 times. Appendix Figure A.4 shows that the positive autocorrelation in physician decisions disappears in the placebo analysis. Within each 4-hour timeframe, environmental factors are likely to be fixed at ED level; however, the autocorrelation in physician decisions disappears as we reassign patients among physicians. This suggests that our results are not driven by common environmental factors at the ED.

In summary, the analyses above suggest that our results remain robust after addressing

[16] The result remains essentially unchanged when we use alternative measures of ED crowdedness: (i) patient waiting time from triage to seeing a physician and (ii) physician-adjusted measure of waiting patients for the wait between triage and seeing a physician.

[17] We divide each day into six periods (0–4, 4–8, 8–12, 12–16, 16–20, and 20–24), and reassign patients starting consultation in each period among on-duty physicians. Placebo test results are robust if we use 2–hour or 6–hour intervals to group patients.

the potential issues relating to common environmental factors. The positive autocorrelation in physician decisions cannot be attributed to common environmental factors.

Additional Results. We conduct more robustness checks in Appendix A.2. We show that the estimated positive autocorrelation in physician decisions is robust to the inclusion of physician multitasking and end-of-shift effects.

4.3 Heterogeneity Analyses

This section examines the heterogeneity of the sequential effect in physician decisions. We find that the estimated positive autocorrelation is higher when the condition of the current patient is associated with greater clinical uncertainty and among less experienced or more fatigued physicians.

Clinical Uncertainty. It has been suggested that sequential effects are more salient when the decision maker faces "ambiguous" situations (Herr, Sherman, and Fazio, 1983; Akaishi et al., 2014). In the context of medicine, conditions with little variation in clinical practices are less clinically ambiguous compared with high-variation conditions (Sabbatini, Nallamothu, and Kocher, 2014). Thus we can use variation in clinical practices to measure clinical uncertainty. Below, we adopt three strategies to examine the relationship between autocorrelation in physician decisions and the clinical uncertainty associated with the current patient.

We first use condition-specific variations in ED admission practices measured in the medical literature. Sabbatini, Nallamothu, and Kocher (2014) study the 15 most common conditions for inpatient admission in the US.[18] For each condition, they compute the variation in admission rates across EDs. We classify the 15 conditions into two groups by the variation in admission rates. Low-variation conditions are the seven conditions with variations less than the median level for the 15 conditions. From our ED data,

[18] Sabbatini, Nallamothu, and Kocher (2014) identify the 15 most frequently admitted conditions based on the 2010 Nationwide Emergency Department Sample in the US: chest pain, soft tissue infections, asthma, COPD, urinary tract infections, fluid and electrolyte disorders, biliary tract disease, cardiac dysrhythmias, diabetes with complications, pneumonia, congestive heart failure, stroke, acute renal failure, acute myocardial infarction, and sepsis. Conditions are presented in descending order from most to least variable in admission practices across EDs.

we extract patients with the 15 most common conditions. Based on this sample, we estimate Equation (1) by adding an indicator for the low-variation condition and its interaction with the lagged admission. Column (1) in Table 4 presents the estimation result. The coefficient on the lagged admission is positive and statistically significant. The coefficient on the interaction term is negative and statistically significant. The degree of positive autocorrelation is significantly smaller for conditions with low variation in admission practices.

Our second measure of practice variations is admission probabilities. We predict admission probability for each patient using a logistic model, with admission decision as the dependent variable and patient characteristics, physician fixed effects, and time fixed effects as explanatory variables. Admission probability closer to 0.5 would indicate a higher degree of variation in admission decisions, while probability closer to either 0 or 1 would indicate lower variation. We divide all patients into three groups by admission probability. The first group includes patients with admission probability less than 0.25, the second 0.25 to 0.75, and the third above 0.75. The proportions of patients who belong to the three groups are 72%, 21%, and 7%. We include group indicators and their respective interactions with the lagged admission in Equation (1), where patients in the second group serve as the reference category. Column (2) in Table 4 presents the estimation results. The coefficient on the lagged admission is positive and statistically significant, and the coefficients on both interaction terms are negative and statistically significant. That is, the positive autocorrelation is larger for patients with moderate admission probabilities than for patients with lower or higher admission probabilities.

Finally, clinical uncertainty can partially be resolved by careful history-taking and physical examination and the use of advanced medical technologies. We examine whether the degree of autocorrelation varies with the length of consultation and with the use of advanced diagnostic imaging—ultrasound, computed tomography scan, and magnetic resonance imaging—for the current patient. Columns (3) and (4) in Table 4 report the estimation results. In column (3), we add consultation length and its interaction with the lagged admission to Equation (1). Column (4) includes the indicator for the use of advanced diagnostic imaging and its interaction with the lagged admission. In both columns, coefficients on the interaction terms are negative and statistically significant at the 1% level. These results suggest that a longer consultation and the use of advanced diagnostic imaging reduce positive autocorrelation in physicians' disposition decisions.

To summarize, the three sets of analyses provide consistent evidence that clinical uncertainty amplifies the degree of positive autocorrelation. Physicians' current decisions are more likely to be influenced by their previous decisions when the condition of the current patient is associated with larger variation in practice. A longer consultation and the use of advanced diagnostic technology reduce clinical uncertainty, and thereby lower the degree of positive autocorrelation in physician decisions.

Physician Experience and Fatigue. We examine whether physician experience and fatigue moderate the degree of positive autocorrelation in admission decisions by including measures of physician experience or fatigue and its interaction with the lagged admission in Equation (1). We define a physician as experienced if her medical experience is at least 7 years—the average value in the sample.[19] Column (1) of Table 5 shows that the degree of positive autocorrelation is significantly higher among less experienced physicians, indicating that medical experience mitigates the degree of positive autocorrelation in physician decisions.

In column (2), we examine physician fatigue, which is measured by the number of hours the physician has worked before seeing the current patient in the shift. Both estimates—on the lagged admission and the interaction term—are positive and statistically significant. The result implies that physicians' disposition decisions are positively autocorrelated at the beginning of the shift, and the degree of autocorrelation increases as time goes by. On average, we find that the degree of autocorrelation increases by approximately 30% in the last hour than in the first hour of an 8-hour shift.

We use an alternative measure for physician fatigue: The length of the rest period before starting the current shift. After a longer rest from the previous shift, physicians are better replenished and less fatigued when starting the current shift. In column (3), the

[19]Physician experience is defined as the number of years since a physician obtained her first degree to practice medicine.

estimate on the interaction is significantly negative. The results in both column (2) and (3) consistently show that the degree of positive autocorrelation increases with physician fatigue.

4.4 Consequences of Autocorrelation

We observe robust positive autocorrelation in physician decisions whereby physicians exhibit a strong tendency to repeat the decision they made for the previous patient. How does this affect the quality of physician decisions? The medical literature suggests that physicians rely on cognitive shortcuts to make fast decisions under uncertainty and time constraints (Groopman, 2007). While these shortcuts can be useful by saving time and resources, they may occasionally lead to increased medical expenditure and adverse patient outcomes (Croskerry, 2002).

We first examine the effect of path dependency on the length of consultation. We define *repeat* as a dummy variable that equals 1 if the physician made the same disposition decision for the current and previous patient and 0 otherwise. We regress the current patient's length of consultation on *repeat*, conditional on patient demographics, triage severity, diagnosis, physician fixed effects, and time fixed effects.[20] Table 6 reports the estimation results with different sets of controls across columns. The estimates on *repeat* are consistently negative and statistically significant. Specifically, column (3) indicates that consultation length is shortened by 15% when the physician repeats her decision. The results are consistent with the notion that physicians' tendency to repeat their decisions helps save time in the ED. This echoes findings in the literature whereby cognitive shortcuts may help save time for physicians (Croskerry, 2002).

Next, we examine the effect of path dependency on patients' medical spending and patient outcomes. Because disposition decisions are positively autocorrelated, the probability of admitting (discharging) the current patient is higher if the physician admitted (discharged) the previous one. This may induce two types of inappropriate dispositions. One is inappropriate discharge: After discharging the previous patient, the physician may

[20] We also control for the disposition mode for the current patient in one specification as a robustness analysis.

be more likely to inappropriately discharge the current patient, who requires inpatient care. The other is inappropriate admission: After admitting the previous patient, the physician may be more likely to inappropriately admit the current patient who does not need inpatient care. While inappropriate admissions increase unnecessary medical spending, inappropriate discharges would escalate medical risk.

We follow Singh (2021) to construct "unexpected" dispositions to proxy for inappropriate dispositions in three steps. First, we predict admission probability for each patient based on patient demographics, triage severity, diagnosis, time fixed effects, and physician fixed effects using a logistic model. We then classify patients who are in the top 22% of the admission probability as "expected" to be admitted and the remaining as "expected" to be discharged.[21] Finally, we define unexpected admission (unexpected discharge) as a dummy variable that equals 1 if the physician admits (discharges) the patient who is expected to be discharged (admitted), and 0 otherwise.

Following Singh (2021), we assume that inappropriate dispositions correlate with our constructed unexpected dispositions. We now examine this assumption. The results reported in the first 4 columns of Table 7 show that our constructed measures of unexpected dispositions capture inappropriate dispositions to some extent. On the one hand, columns (1) and (2) show that unexpected admission is not correlated with subsequent admission within 15 days and ambulance use within 15 days after the current visit.[22] This is consistent with what we expect to observe—an inappropriate admission increases medical costs but it should have limited impact on patient health outcomes. On the other hand, columns (3) and (4) show that unexpected discharge has downstream health consequences for the patient: The probability of subsequent admission increases by 1.5 percentage points and the probability of subsequent ambulance use increases by 0.5 percentage point. Both estimates are statistically significant at the 1% level.

Columns (5) and (6) show that a lagged admission decision increases the probability of unexpected admission for the current patient by 1.3 percentage points and decreases the

[21] In the full sample, 22% of patients are admitted.

[22] It is hard to fully measure the appropriateness of disposition decisions. We only observe inpatient admission through and ambulance to this specific ED in our data. We do not have other measures for patient health outcomes given our ED data.

probability of unexpected discharge by 1.5 percentage points. An equivalent interpretation of this result is that a lagged discharge decision decreases unexpected admission by 1.3 percentage points and increases unexpected discharge by 1.5 percentage points. In our sample, 78% of patients are discharged, which implies that most patients are treated after a lagged discharge decision. Hence, overall, the positive autocorrelation leads to an increase in unexpected discharge for high-risk patients and a decrease in unexpected admission for low-risk patients.[23] Furthermore, Appendix Table A5 indicates that for a high-risk patient, a lagged discharge decision leads to a 4.2-percentage-point increase in the probability of unexpected discharge, while for a low-risk patient, a lagged admission decision results in a 2.7-percentage-point increase in the probability of unexpected admission.

5 Sequential Decisions in Obstetrics

5.1 Institutional Setting and Administrative Data

Institutional Setting. Our second empirical setting concerns physicians' decisions for childbirth. With almost 4 million babies born each year, childbirth is the most common cause for hospital admissions; also, C-sections are the most common inpatient surgery in the US. Childbirth can be performed vaginally or by C-section. A C-section is a major abdominal surgery intended for high-risk childbirths in which a vaginal delivery would put the baby or the mother at risk. Both the World Health Organization (WHO) and the American College of Obstetricians and Gynecologists recommend that C-sections be performed only when medically necessary (Caughey et al., 2014; Betrán et al., 2016). C-sections are associated with an overall increase in poor outcomes for pregnancies that are not high risk (Caughey et al., 2014). They are also more expensive, require longer hospital stays, and have slower recoveries.

In the US, the decision to perform a C-section is typically made by the physician.

[23] By our definitions above, the value of the dummy variable unexpected discharge equals 1 only for discharged patients whose predicted risk is among the top 22% in the sample. Similarly, the value of the dummy variable unexpected admission equals 1 only for admitted patients whose predicted risk is among the bottom 78% in the sample. Therefore, by construction, unexpected discharge (admission) is only relevant for high-risk (low-risk) patients.

Several patient conditions increase the probability of a C-section. For example, about 90% of mothers who received a C-section in the past receive a C-section again, even though recent literature suggests that many may benefit from attempting vaginal birth after a C-section (Caughey et al., 2014). Another strong predictor of a C-section is breech birth, in which the baby is positioned head-up in the uterus. Other conditions that increase the likelihood of receiving a C-section include disproportion, multiple births, and problems with the placenta or umbilical cord.

For high-risk mothers, physicians will often schedule a C-section ahead of time. The scheduling typically occurs after the 39th week of gestation to minimize the risk of complications. Otherwise, the decision to perform a C-section occurs during labor. As labor progresses, the physician must trade off the risk of allowing labor to continue against the risk of performing a C-section.[24]

Administrative Data. We use administrative hospital discharge data on all births that occurred in New York State from January 1, 2005, through December 31, 2015. The data are provided by the Statewide Planning and Research Cooperative System of New York and include comprehensive records for each admission, including patient characteristics, physician identifier, clinical decisions, and procedural timestamps.

Over the 11-year period, there were over 2.5 million births in the State of New York. We limit our attention to physicians who performed at least 100 deliveries during the time period. The resulting analytic sample size is 2,458,773 deliveries performed by 3,105 physicians. Appendix Table B.2 reports summary statistics for the analytic sample. The average C-section rate is 33.8%. The mothers' average age is 29 years. The proportion by race is 51% White, 18% African American, 6% Asian, 1% Native American or Alaskan Native, and 25% other races. The remainder of the table shows summary statistics of the 11 patient conditions that are included as controls in our regression analyses.

[24] Long or difficult vaginal deliveries can increase the risk of fetal trauma (Baskett et al., 2007). Vaginal deliveries also carry higher risk of perineal lacerations and pelvic floor damage, which can lead to sexual dysfunction and fecal and urinary incontinence (Fenner et al., 2003).

5.2 Comparison Between the ED and Childbirth Settings

Institutional Contexts. Several important institutional differences between the ED and childbirth settings render childbirth an interesting secondary setting for the study of sequential physician decisions. First, whereas ED physicians are generalists who treat patients with a wide range of medical conditions, obstetricians are specialists in pregnancy and childbirth. Second, in contrast to the frequent decisions of ED physicians, obstetricians make delivery decisions less frequently in succession.[25] Finally, pregnancies are carefully observed and monitored over the course of prenatal visits, and as a result physicians tend to be better informed when making treatment decisions. We test whether, despite these key differences, positive autocorrelation also exists in obstetricians' delivery decisions. Consistent findings between the ED and childbirth settings would suggest a potentially large scope of sequential effects in physician decisions.

Data. The obstetric data provide a detailed account of childbirth on a large scale. The size and representativeness of the sample distinguish these data from the ED data examined above. Whereas the ED data include all patient visits in 2 years from a single hospital, the obstetric data record all births across hospitals in New York State for an 11-year horizon. The secondary obstetric data therefore allow us to test the external validity of our primary findings from the ED. On the other hand, records from the ED data are more granular and include information on medical resource utilization and other medical staff, which cannot be obtained from the obstetric data.

Identifying Assumptions. Our first identifying assumption of the conditionally random ordering of patients is plausible in the childbirth setting, since the timing of when labor begins is unexpected for most deliveries.[26] As in the ED setting, we perform nonparametric

[25]In our dataset, the obstetrician's previous delivery was on the same day as the current delivery 34% of the time; 1 day earlier 23% of the time; 2 days earlier 10% of the time; and 3 or more days earlier 34% of the time.

[26]The two exceptions are scheduled C-sections and induced labor. C-sections are often scheduled ahead of time for high-risk mothers. The physician may also schedule an induction of labor if the delivery is overdue or complications exist, or at the request of the patient. Some mothers who start with induced labor end up receiving a C-section. In a recent study, 27% of induced labor resulted in a C-section (Davey and King, 2016). In robustness analysis, we conduct the analysis with a subsample of weekend deliveries that are unlikely to be scheduled ahead of time. The results remain highly significant and robust.

runs tests; the results suggest that patient sequence follows a random order with respect to patient age and race (Appendix B.1). In our robustness analysis, we show that the positive autocorrelation in C-sections persists when we rule out plausibly scheduled deliveries.

Testing the second identifying assumption, however, is constrained by the limited information available in the obstetric data. The data do not provide any information on hospital resource capacity or other medical staff who assist the physician during childbirth. Hence, it is infeasible to verify whether our results are confounded by common environmental factors in the childbirth setting.

With these differences in mind, the obstetrics context provides an important secondary setting to complement our analysis of path dependency in physician decisions in the ED. However, given the data limitations, caution is warranted in interpreting the positive autocorrelation in C-sections as a causal effect, since we are not able to address potential issues relating to common environmental factors.

5.3 Empirical Analyses

Empirical Specification Details. We estimate Equation (1) to test autocorrelation in physician's delivery decisions. Y_{it} is an indicator for whether the physician performs a C-section on the current delivery and $Y_{i,t-1}$ is an indicator for whether the physician performed a C-section on the previous delivery. The ordering of patients to physician is determined by the time of admission, t. We control for patient demographics (age and race),[27] medical conditions, time fixed effects, and physician fixed effects unless stated otherwise. Medical conditions include an indicator for the mother's having a history of previous C-section, placenta previa, disproportion, breech, twin (or multiple) birth, preeclampsia, hypertension, diabetes, early labor, late labor, and long labor.

Basic Results. Panel A of Table 8 reports fixed-effect estimates from Equation (1). Column (1) reports results for the full sample. The estimate of α_1 is positive and statistically significant at the 1% level. The probability of performing a C-section is 1.3 percent-

[27]Different from the ED setting, patients are divided into six groups by age in the childbirth setting: <20, 20–24, 25–29, 30–34, 35–39, and 40 or above.

age points higher when the physician's previous delivery was a C-section, compared with when the physician's previous delivery was a vaginal birth. The estimate represents a 4.0% increase relative to the mean C-section rate of 0.338, which suggests strong positive autocorrelation in delivery decisions.

Panel B of Table 8 shows pooled OLS estimates without controlling for physician fixed effects. Across the board, the pooled OLS estimates are larger than the fixed-effect estimates. For the full sample in column (1), taking the fixed-effect estimate as the lower bound and the pooled OLS estimates as upper bounds, the physician is 4.0%–12.9% more likely to perform a C-section after having recently performed a C-section. The effect size is comparable to that in Bindler and Hjalmarsson (2019), who find that in jury decisions, a lagged guilty verdict increases the chance of a subsequent guilty verdict by 6.7%–14.1%.

Consistent with our findings from the ED setting, we find significantly positive autocorrelation in physician decisions for childbirth. The magnitude of the autocorrelation is smaller in the childbirth setting than the ED setting. This difference could be explained by the longer time gap—sometimes 1, 2, or 3 days—between consecutive deliveries for obstetricians, which may weaken the impact of the previous decision on the current delivery decision. It could also be because obstetricians usually possess more information about their patients. In column (2), we restrict the sample to deliveries in which the physician's previous delivery was within 24 hours. The magnitude of the autocorrelation doubles to 0.028, which implies that the likelihood of a C-section increases by 2.8 percentage points after a previous C-section on the same day. This is equivalent to an 8.8% increase in the C-section rate from the mean, and suggests that the magnitude of autocorrelation increases if the previous delivery decision is more recent.

Robustness Analyses: Conditional Random Ordering. A potential concern for our identification is that some deliveries—namely, scheduled C-sections and induced labor—are scheduled ahead of time, and the scheduling of these deliveries may be endogenous to the decision. For example, if C-sections tend to be scheduled close together, this would create a spurious positive autocorrelation in our results. Due to data limitations, we are unable to identify which deliveries were scheduled ahead of time. We address this potential concern in two ways. First, in column (3) we restrict the sample to weekend deliveries, which are almost always emergency deliveries and thus patient ordering is likely to be random. The estimated coefficient is positive and statistically significant at the 1% level. Furthermore, the magnitude of the coefficient in column (3) is comparable to that in column (1).

Second, in column (4) we remove breech births and mothers who have had a previous C-section; these are the two most common conditions in which C-sections are scheduled ahead of time.[28] The estimated coefficient in column (4) is 0.014, which is almost identical to the estimated coefficient in column (1). The magnitudes of the estimated coefficients remain almost unchanged across columns (1), (3), and (4), which suggests that the potential endogenous scheduling of C-sections may not be a major concern.

Additional Analyses. We examine the heterogeneity of our findings in Appendix B.2. Consistent with our findings from the ED setting, we find that the degree of autocorrelation in C-section decisions increases when the current patient is associated with larger variation in C-section risk. However, we are not able to examine the role of physician experience and fatigue in the childbirth setting, due to the lack of information in the obstetric data.

6 Mechanism

This section discusses potential mechanisms underlying the observed positive autocorrelation in physician decisions. We fist examine memory and attention-based anchoring and adjustment, as proposed by Bordalo, Gennaioli, and Shleifer (2020). We find that the model can account for the empirical patterns we observed. We then present additional evidence from both reduced-form and structural estimates, which provides further support for the model. Finally, we discuss alternative mechanisms, including quotas and learning, and show that while these mechanisms may contribute to some of the observations, they

[28] These are the top two predictors of receiving a C-section in our data: 90.2% of breech births receive a C-section and 87.1% of births in which the mother has a history of receiving a C-section before receiving another C-section. Physicians often schedule C-sections for these patients.

can not fully explain our findings.

6.1 Memory and Attention Mechanism

Bordalo, Gennaioli, and Shleifer (2020) propose a theoretical framework to model the memory and attention mechanism in decision-making, and show that the model provides a unified account for a range of behavioral anomalies documented in the literature. In their framework, a choice option cues the attention to and recall of the individual's past experiences from the memory database, which are weighted by their similarity to the cue and are then consolidated into a norm. The model explains positively and negatively auto-correlated decisions simultaneously: Positive autocorrelation arises when the discrepancy between the current stimulus and the memory and attention-based norm is moderate, and negative autocorrelation prevails when the discrepancy is unexpectedly large. Based on this mechanism, the model can also account for the varying degrees of autocorrelation in sequential decisions when the current choice environment is compared with past experiences and a higher weight is assigned to more recent ones. Below we apply the model of Bordalo, Gennaioli, and Shleifer (2020) to the context of physician decisions.

Consider a physician who treats a total of T patients within a shift and label each patient by time period $t \in \{1, ..., T\}$. We refer to the physician as "she" and the patient as "he." Let C be the set of patients' characteristics that can be observed by the physician, including age, gender, disease category, triage severity, etc., and $c_t \in C$ be the characteristics of patient t. Let $r : C \to R$ be the function that maps the patient's observed characteristics to his clinical risk and $r_t = r(c_t)$ be the clinical risk of patient t.

We assume that the physician forms a norm (r_t^n) to assess the risk of patient t ($t \geq 2$) by retrieving experiences from both her recent and long-term memories. The physician's recent memory is based on the treatment of all patients prior to the current one within the same shift; we denote by r_{t-k} the clinical risk of the patient lagged k periods before the current patient. Her long-term memory is based on the treatment of all patients prior to the current shift; we denote by r_t^m the average risk of patients recalled by the physician (based on the cue from patient t) from her long-term memory. The physician's norm r_t^n

is a weighted average of $\{r_{t-k}\}_{k=1}^{t-1}$ and r_t^m:

$$r_t^n = \frac{\sum_{k=1}^{t-1} \alpha e^{-\beta(k-1)} S(c_{t-k}, c_t) r_{t-k} + r_t^m}{\sum_{k=1}^{t-1} \alpha e^{-\beta(k-1)} S(c_{t-k}, c_t) + 1}, \tag{2}$$

where the weight of r_t^m is normalized to one. The weight of r_{t-k} is the product of three terms: (i) $\alpha(> 0)$ captures the relative importance of the recent memory to the long-term memory; (ii) $e^{-\beta(k-1)}$ captures the similarity between patient $t-k$ and t in the time domain, where $\beta > 0$; and (iii) $S(c_{t-k}, c_t)(\geq 1)$ captures the similarity between patient $t-k$ and t in their observable characteristics. We assume that similar experiences are more likely to be recalled and receive more attention from the physician; that is, $S(c_{t-k}, c_t)$ increases as c_{t-k} and c_t become more similar.

The norm r_t^n serves as the physician's anchor when assessing the risk of patient t. She adjusts her perceived risk (r_t^p) from the anchor based on her perceived difference between r_t and the anchor r_t^n in the following manner:

$$r_t^p = r_t^n + \sigma(r_t, r_t^n)(r_t - r_t^n), \tag{3}$$

where $\sigma(r_t, r_t^n)(r_t - r_t^n)$ is the physician's perceived difference, and the term $\sigma(r_t, r_t^n)$ captures the salience of the difference between r_t and r_t^n; $\sigma(r_t, r_t^n) = 0$ when $r_t = r_t^n$.[29]

The function $\sigma(r_t, r_t^n)$ is smaller than 1 when the difference between r_t and r_t^n is small; it is greater than 1 when the difference is large. In the former case, the physician's perceived risk for the patient, r_t^p, is anchored to the norm r_t^n,[30] resulting in positive association between r_t^p and r_t^n. This is referred to as anchoring. In the latter case, the physician is surprised by the difference and overadjusts the risk against the norm; consequently, r_t^p is adjusted away from the norm r_t^n,[31] leading to negative association between r_t^p and r_t^n. This is referred to as adjustment.

[29]See Bordalo, Gennaioli, and Shleifer (2012, 2013) for how salience affects choices and Bordalo, Gennaioli, and Shleifer (2022) for a survey on salience.

[30]That is, $r_t^n > r_t^p > r_t$ when $r_t^n > r_t$, and $r_t^n < r_t^p < r_t$ when $r_t^n < r_t$.

[31]That is, $r_t > r_t^p$ when $r_t^n > r_t$, and $r_t < r_t^p$ when $r_t^n < r_t$.

Specifically,

$$\frac{\partial r_t^p}{\partial r_t^n} = 1 - \sigma(r_t, r_t^n) + \frac{\partial \sigma(r_t, r_t^n)}{\partial r_t^n}(r_t - r_t^n). \tag{4}$$

When r_t is close to r_t^n, both $\sigma(r_t, r_t^n)$ and $\frac{\partial \sigma(r_t, r_t^n)}{\partial r_t^n}(r_t - r_t^n)$ are close to 0; hence, $\frac{\partial r_t^p}{\partial r_t^n} > 0$. By contrast, when r_t is far from r_t^n, the salience term $\sigma(r_t, r_t^n)$ is larger than one, and $\frac{\partial r_t^p}{\partial r_t^n} < 0$.[32]

This model, with the assumption of recency effect, accommodates both positive and negative autocorrelation in sequential medical decisions. Specifically, a physician with recency effect pays more attention to the previous patient in her memory database, so the previous patient's risk r_{t-1} is disproportionately weighted higher in the risk norm r_t^n (Equation (2)). The aforementioned association between r_t^p and r_t^n would then lead to autocorrelation between r_t^p and r_{t-1}. That is, when the difference between r_{t-1} and r_t is small, the previous patient's risk r_{t-1} increases the perceived risk r_t^p for patient t—and consequently positive autocorrelation occurs, since the probability of admission monotonically increases with the perceived risk. Conversely, when the difference is unexpectedly large, negative autocorrelation emerges.[33]

To sum, the model helps explain our observed positive autocorrelation and gives rise to three additional predictions: (i) the recency effect, whereby the degree of autocorrelation decreases with the time distance between the current and earlier patients within the same shift; (ii) the similarity effect, whereby the degree of autocorrelation increases with the similarity in characteristics between the current and previous patients; and (iii) potential negative autocorrelation—whereas positive autocorrelation is prevalent in physician decisions, negative autocorrelation can occur when the physician is surprised by a large difference between the current and previous patient and overadjusts the risk against the norm. We next examine these three predictions.

[32] The assumption that guarantees the inequality is as follows: The function $\sigma(r_t, r_t^n)$ increases with the difference between r_t and r_t^n when the sign of $r_t - r_t^n$ remains unchanged. It follows that $\frac{\partial \sigma(r_t, r_t^n)}{\partial r_t^n}(r_t - r_t^n) \leq 0$, and thus $\frac{\partial r_t^p}{\partial r_t^n} < 0$.

[33] The anchoring and adjustment mechanism also accounts for the observation in Section 4.3—the positive autocorrelation is higher when the current patient is of larger clinical uncertainty. More specifically, patients with lower clinical uncertainty are likely to have a condition that is either very mild or very severe. As a result, they are less likely to be similar to the previous patient in terms of risk, and therefore, we are less likely to observe positive autocorrelation.

6.2 Reduced-form Estimates

Recency Effect. Equation (2) shows that the weight term $e^{-\beta(k-1)}$ decreases over time as memory fades. To test this recency effect, we examine the extent to which physicians' treatment decisions are influenced by more distant decisions in the past. In both the ED and childbirth settings, we find significant diminishing effects over longer lagged decisions (Appendix Tables A.6 and B.3). In the ED setting, the likelihood of admitting the current patient increased by 2.5 percentage points after admitting the previous patient, compared with 1.0, 0.9 and 0.8 percentage-point increases after admitting the patients 2, 3 and 4 time periods prior, respectively.[34] Similarly, in the childbirth setting, having performed a C-section on the previous patient was associated with a 1.3 percentage-point increase in C-section risk, whereas having performed a C-section on patients 2, 3 and 4 time periods ago was associated with 0.5, 0.5 and 0.6 percentage-point increases in C-section risk, respectively.

Similarity Effect. Equation (2) shows that the weight assigned to the previous patient depends not only on recency but also on the similarity in patient characteristics. In both the ED and childbirth settings, we find evidence consistent with the insights arising from the model: The degree of autocorrelation increases when consecutive patients are more similar in their characteristics (Appendix Tables A.7 and B.1). In the ED, on average, the degree of autocorrelation increases by more than 200% when the two patients are similar in triage severity and by around 50% when the two patients are similar in disease or age.[35] By contrast, the degree of autocorrelation does not significantly vary by similarities in race or gender. Given that race and gender are less relevant predictors of and not significantly correlated with clinical risk in the ED, the results suggest that physicians, while relying on shortcuts, selectively retrieve memories based on similarities in the characteristics relevant to clinical risk and not on similarities in other characteristics.

[34] We also find that in the ED setting, the estimated autocorrelation decreases when the physician takes a break before attending to the current patient.

[35] Similarity in triage severity implies that the two patients are both severe or both non-severe. Similarity in disease implies that the two patients belong to the same category out of the 18 broad diagnostic categories. Similarity in age implies that the age gap between the two patients is within 6 years (the first quartile of age gaps); results are robust if we use alternative age gaps—e.g., 3, 5, or 10 years.

Potential Existence of Negative Autocorrelation. To investigate the potential existence of negative autocorrelation in the ED, we divide consecutive patient pairs into 8 groups by the three dimensions relevant to clinical risk—severity, disease, and age. We examine heterogeneous autocorrelation across the 8 groups by estimating the following the equation:

$$Y_{it} = \rho_0 + \sum_{j=1}^{8} \beta_j Y_{i,t-1} \times \mathbf{1}(G_{it} = j) + G_{it}\pi + X_{it}\gamma + \delta_i + \eta_{it}, \tag{5}$$

where G_{it} is a vector of group indicators; other variables are the same as in Equation (1). Since $Y_{i,t-1}$ is omitted from the right-hand side of Equation (5), the 8 coefficients of β_j capture the degree of autocorrelation across the 8 groups.

Columns (1) and (2) of Table 9 present the pooled OLS and fixed-effect estimates for Equation (5), respectively. Estimates are similar between the two columns. The results show that *positive autocorrelation* prevails, since the estimates of β_j are positive in 7 out of the 8 groups in each column. Moreover, the magnitude of the estimates generally increases when consecutive patients are similar in more dimensions of characteristics. For example, the estimate of β_1 is positive but small in magnitude, which suggests weakly positive autocorrelation when the two patients are similar in none of the three dimensions. By contrast, the estimate of β_8 is positive, large in magnitude, and statistically significant at the 1% level, which suggests that the physician is more likely to anchor on the previous patient when the two patients are similar in all of the three dimensions.

Table 9 also documents suggestive evidence of *negative autocorrelation* in physician decisions. The estimate of β_4 is negative—namely, the physician's decisions appear to be negatively autocorrelated when the previous and current patients are similar in disease and age but not in triage severity. In our setting, triage severity is one of the most important determinants for patient disposition; the admission rate is 56% in severe cases and 8% in non-severe cases. The similarities in age and disease might cue the physician to expect the current patient to have medical conditions similar to the previous one, and thus use the previous patient as an anchor. Yet the difference in triage severity could surprise the physician to contrast the current patient with the previous one, so she overadjusts her perceived risk for the current patient away from the anchor, which leads to negative autocorrelation in physician decisions. Some caution should also be exercised in interpreting this finding, since the estimate is only marginally significant at the 10% level.

Discussion. In summary, we document reduced-form evidence consistent with the memory and attention-based anchoring and adjustment mechanism. Physician decisions are generally positively autocorrelated and the degree of autocorrelation is larger when the previous and current patient are closer in time (recency effect) or more similar in their observable characteristics (similarity effect). We also find some suggestive evidence that physicians may make negatively autocorrelated decisions in the face of a surprise.

Overall, our results show that positive autocorrelation prevails in healthcare settings. We suggest two reasons. First, physicians, as medical experts, have accumulated extensive experience from years of training and working with a wide variety of patients.[36] Physicians would be able to retrieve cases similar to patient t from their long-term memory, such that r_t^m in Equation (2) is close to r_t. When their norms are largely determined by those retrieved from long-term memory and, to a lesser extent, by their recently treated patients in the shift, α is small in Equation (2). As a result, physicians' norms would be close to the current patient's risk. That is, the gap between r_t^n and r_t is small in Equation (3), the salience term is less than one, and positive autocorrelation prevails.[37]

Second, we observe physicians' binary decisions in our datasets but not risk ratings, so it is hard to detect negative autocorrelation from binary choice data. The reason is as follows. The necessary condition for negative autocorrelation is that the gap between the current patient's risk r_t and the norm r_t^n is sufficiently large. Equation (2) shows that the norm r_t^n is a weighted average between the risk levels of recently treated patients

[36] For example, in the U.S., predetermined qualifications for physician licensure includes medical school graduation, residency training, and passing a comprehensive national medical licensing examination, which in total may take a decade or longer (Bhattacharya, Hyde, and Tu, 2018).

[37] Negative autocorrelation may prevail in some settings. For example, using large-scale data on study grant admissions and job hiring interviews, Radbruch and Schiprowski (2022) show that the assessment of a candidate decreases in the quality of other candidates assigned to the same evaluator, and provide empirical support for negatively autocorrelated decisions caused by the interplay between the associative recall of prior candidates and the attention to salient quality differences.

within the same shift ($\{r_{t-k}\}$) and those retrieved from the long-term memory database (r_t^m). When physicians are able to retrieve cases similar to patient t from their long-term memory, r_t^m is similar to r_t. To observe negatively autocorrelated decisions, we thus need r_t to dramatically differ from $\{r_{t-k}\}$. This condition is satisfied in only two scenarios when considering disposition decisions in the ED setting. First, previous patients have low risk, but the current patient has high risk and thus should be admitted with almost certainty; second, previous patients have high risk, but the current patient has low risk and thus should be discharged with almost certainty. The large difference between the two consecutive patients would push the physician's perceived risk for the current patient away from the norm—i.e., the physician would perceive the risk of the current patient to be even higher in the first scenario and lower in the second scenario. In both scenarios, however, the physician's binary disposition decision for the current patient would hardly be affected, since the current patient should be admitted or discharged with almost certainty.

6.3 Structural Estimates

Next, we structurally estimate the model presented in Section 6.1. The structural estimation aims to help better understand the memory and attention mechanism underlying the observed positive and negative autocorrelation in physician decisions in Table 9.

Parameterization. We first parameterize the risk norm function in Equation (2) in three steps. In the first step, for the physician's recent experiences, we consider only the risk of the most recent patient (r_{t-1}) among all patients within the same shift ($\{r_{t-k}\}$). This simplification is in part due to the observed recency effect, whereby the decision for the most recent patient has a significantly larger effect on the current decision compared with decisions for patients lagged 2 or more periods (Appendix Table A.6). In the second step, we use r_t as a proxy of r_t^m. As discussed in the section above, physicians have accumulated extensive experience during their years of training and working. Thus, they are able to retrieve patients who are similar to the current patient from their long-term memory databases. In the third step, we specify the similarity function as

$$S(c_t, c_{t-1}) = \lambda_{\text{SEV}}^{I_{t,t-1}} \lambda_{\text{DZ}}^{J_{t,t-1}} \lambda_{\text{AGE}}^{K_{t,t-1}}, \tag{6}$$

where the dummy indicators $I_{t,t-1}$, $J_{t,t-1}$, and $K_{t,t-1}$ indicate whether patient t and $t-1$ are similar in severity, disease, and age, respectively. In the similarity function, λ_{SEV} is the parameter of similarity in severity between the current and previous patients, which measures the ratio of the weight of the previous patient's risk r_{t-1} when the two patients are similar in severity over the weight when they are not; λ_{DZ} and λ_{AGE} are the parameters of similarity in disease and age, respectively. These three parameters will be estimated. Combining the three steps, the risk norm function is

$$r_t^n = \frac{\lambda_{\text{SEV}}^{I_{t,t-1}} \lambda_{\text{DZ}}^{J_{t,t-1}} \lambda_{\text{AGE}}^{K_{t,t-1}} \alpha r_{t-1} + r_t}{\lambda_{\text{SEV}}^{I_{t,t-1}} \lambda_{\text{DZ}}^{J_{t,t-1}} \lambda_{\text{AGE}}^{K_{t,t-1}} \alpha + 1}. \tag{7}$$

Second, we specify the salience term as

$$\sigma(r_t, r_t^n) = \sigma \cdot \frac{|r_t - r_t^n|}{\bar{r} - \underline{r}}, \tag{8}$$

where $\sigma(> 0)$ captures how the physician perceives the difference between the norm and the current patient's risk as salient. The two numbers $\bar{r}$ and $\underline{r}$ represent the maximal and minimal risk over all patients in the sample, respectively. This simple functional form maintains the key features discussed in Section 6.1. First, $\sigma(r_t, r_t^n)$ increases with the difference between the norm r_t^n and the current patient's risk r_t. Second, the reaction of the physician's perceived risk to a change in the norm reduces to

$$\frac{\partial r_t^p}{\partial r_t^n} = 1 - 2\sigma \cdot \frac{|r_t - r_t^n|}{\bar{r} - \underline{r}}, \tag{9}$$

which is positive when r_t is close to r_t^n and negative when they are distinct given a reasonably large σ.

Plugging Equations (7) and (8) into Equation (3), the physician's perceived risk for patient t is given by

$$r_t^p = \frac{\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\alpha r_{t-1} + r_t}{\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\alpha + 1} + \sigma\left(\frac{\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\alpha}{\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\alpha + 1}\right)^2 \frac{|r_t - r_{t-1}|(r_t - r_{t-1})}{\bar{r} - \underline{r}}. \tag{10}$$

Estimation. We first estimate a logistic model of patient admission on patient demographics, triage severity, diagnosis, and time fixed effects. Using the estimated coefficients, we compute r_t, r_{t-1}, $\bar{r}$, and $\underline{r}$.

We then use maximum likelihood estimation to estimate the parameters, λ_{SEV}, λ_{DZ}, λ_{AGE}, α, and σ. We specify that the patient's utility from admitting patient t is $U_{t,1} = r_t^p + \epsilon_t$, where the deterministic part of the utility from admission increases with the physician's perceived risk for the patient, defined in Equation (10); her utility from discharging the patient is $U_{t,0} = \hat{\epsilon}_t$, where the deterministic part of the utility from discharge is normalized as 0. The two error terms ϵ_t and $\hat{\epsilon}_t$ are independent of patients' observable characteristics.[38] The physician admits the patient if and only if $r_t^p + \epsilon_t \geq \hat{\epsilon}_t$. We assume that the two error terms follow type I extreme value distribution. The probability of admitting patient t is $\frac{e^{r_t^p}}{1+e^{r_t^p}}$. We obtain the likelihood function conditional on the observed characteristics of the patients as

$$\mathcal{L}(\alpha, \sigma, \lambda_{\text{SEV}}, \lambda_{\text{DZ}}, \lambda_{\text{AGE}} | \{c_t\}_{t\in\Gamma\cup\Omega}) = \prod_{t\in\Omega}\left(\left(\frac{e^{r_t^p}}{1+e^{r_t^p}}\right)^{A_t}\left(\frac{1}{1+e^{r_t^p}}\right)^{1-A_t}\right).$$

The set Γ contains all patients who are the first in their corresponding shifts, Ω contains those who are not the first in their corresponding shifts, and A_t indicates whether patient t is admitted ($A_t = 1$) or not ($A_t = 0$). In the likelihood function, conditional on all patient characteristics, the physician's sequential decisions are independent since her perceived risk for each patient depends on patient characteristics. The observed autocorrelation in physician decisions is generated by the correlation between the perceived risk of the current patient and the risk of the previous patient, as described by Equation (10).

We have five parameters—λ_{SEV}, λ_{DZ}, λ_{AGE}, α and σ—to be estimated. The three

[38] We assume that ϵ_t and $\hat{\epsilon}_t$ are independent conditional on r_t^p. Physician admission decisions for patient t and $t-1$ are correlated through r_t^p but not through the error terms.

similarity parameters λ_{SEV}, λ_{DZ}, and λ_{AGE} can be separately identified because of the conditional random assignment of patients. We mainly rely on the functional form to separately identify α and σ. Equation (10) shows that the influence of α on r_t^p relies on both r_t, r_{t-1}, and the quadratic function of their difference; by contrast, the influence of σ on r_t^p relies only on the quadratic function of the difference.

Results. Table 10 presents the parameter estimates. Estimates of the similarity parameters $\lambda_{\text{SEV}}, \lambda_{\text{DZ}}$, and λ_{AGE} are all larger than one. This supports the mechanism of similarity-based retrieval: The weight of the previous patient's risk increases when the current and previous patient are similar in their characteristics. In particular, λ_{SEV} is larger than both λ_{DZ} and λ_{AGE}, which indicates that similarity in severity renders a heavier weight on the previous patient's risk in forming the norm, compared with similarities in age and disease. This result is consistent with the reduced-form evidence reported in Appendix Table A.7.

Estimates of the weight on the previous patient's risk α is 0.01 when the previous and current patient are similar in none of the three characteristics. This result is consistent with the estimate of β_1 reported in Table 9. Combining the estimates of α with $\lambda_{\text{SEV}}, \lambda_{\text{DZ}}$, and λ_{AGE}, the weight on the previous patient's risk is 0.09 when the two patients are similar in all of the three characteristics. Since the weight on the risk of the current patient, which is used to proxy for the average risk of patients retrieved from long-term memory, is normalized to 1, the weight on the previous patient's risk is low. This result suggests that the physician is less affected by the previous patient's risk in forming the norm and that the gap between the current patient's risk and the norm is small, which explains the prevalence of positive autocorrelation in our context. Finally, the estimate of salience parameter σ is 64.75. Since the estimate of α is 0.01, the estimate of σ is sufficiently large to account for both positive and negative autocorrelation (Equation (10)).

Discussions. The estimated model accommodates both positive and negative autocorrelation. We take the derivative of Equation (10) with respect to r_{t-1} and plug in the parameter estimates of α and σ:[39]

$$
\begin{aligned}
\frac{\partial r_t^p}{\partial r_{t-1}} &= \frac{\partial r_t^p}{\partial r_t^n}\frac{\partial r_t^n}{\partial r_{t-1}} \\
&= \frac{\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\alpha}{\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\alpha+1}\left(1-\frac{2\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\alpha\sigma}{\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\alpha+1}\frac{|r_t-r_{t-1}|}{\bar{r}-\underline{r}}\right) \\
&\approx 0.01\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\left(1-1.27\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}\frac{|r_t-r_{t-1}|}{\bar{r}-\underline{r}}\right).
\end{aligned}
\tag{11}
$$

The above equation shows that the sign of the autocorrelation equals the sign of the term in parentheses. Here, we observe a trade-off. On the one hand, when the current and previous patients are similar, such that $I_{t,t-1}$, $J_{t,t-1}$, and $K_{t,t-1}$ all equal one, $\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}$ attains its maximal value. At the same time, the risk difference between the current and previous patients ($|r_t-r_{t-1}|$) attains its minimal across the 8 groups we analyzed in Section 6.2. On the other hand, when the current and previous patients are different, such that $I_{t,t-1}$, $J_{t,t-1}$, and $K_{t,t-1}$ all equal zero, $\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}$ attains its minimal value, and the risk difference between the current and previous patient ($|r_t-r_{t-1}|$) attains its maximal. In both cases, $\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}|r_t-r_{t-1}|$ does not achieve its maximal value and the term in parentheses is not likely to be negative; consequently, negative autocorrelation does not occur. To observe negative autocorrelation, we need to balance between $\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}$ and $|r_t-r_{t-1}|$. For example, when the two patients are similar in age and disease ($J_{t,t-1}=1$ and $K_{t,t-1}=1$) but different in severity ($I_{t,t-1}=0$), $\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}$ is larger than one and $|r_t-r_{t-1}|$ is also large, since severity is a major determinant of patient risk. In this case, $\lambda_{\text{SEV}}^{I_{t,t-1}}\lambda_{\text{DZ}}^{J_{t,t-1}}\lambda_{\text{AGE}}^{K_{t,t-1}}|r_t-r_{t-1}|$ may achieve the maximal and negative autocorrelation may occur. This is consistent with the negative estimate of β_4 in Table 9. In this regard, the memory and attention mechanism (Bordalo, Gennaioli, and Shleifer, 2020) explains both the positive and negative estimates of autocorrelation in physician decisions in Table 9.

[39] In the partial derivative, we assume that r_{t-1} changes, holding $I_{t,t-1}$, $J_{t,t-1}$, and $K_{t,t-1}$ constant. This assumption holds in our heterogeneity analysis across the 8 groups in Table 9.

6.4 Alternative Explanations

We discuss quota and learning as two alternative mechanisms that could in principal generate positive autocorrelation in physician decisions.[40]

Quota. Physicians may face a constraint in the total number of affirmative responses (i.e., a quota). For example, the number of admission decisions in the ED is subject to the availability of inpatient beds. As discussed in Section 4.2, the positive autocorrelation in decisions persists after controlling for the level of bed occupancy and crowdedness in the ED, which implies that quota is unlikely to be the underlying mechanism behind the observed autocorrelated decisions of physicians.

Learning. Decision makers may be unsure about how to set the threshold to reach a decision, so they learn over time. For example, an ED physician may update her belief over patients' conditions and adjust her threshold for admission through repeated practice. The learning interpretation is less consistent with the observed strong recency effect: The positive autocorrelation is primarily driven by the immediately preceding decisions, and the effect dissipates rapidly over lagged decisions.

7 Conclusion

This paper documents positive autocorrelation in physician decisions, using large-scale data from two independent healthcare settings: inpatient-admission in the ED and C-section decisions for childbirth. The degree of autocorrelation is higher when consecutive patients share more similar characteristics, when the current patient is of larger clinical uncertainty, and when the physician is less experienced or more fatigued. We also find that decisions can occasionally be negatively autocorrelated in the face of a surprise. Our results are most consistent with the mechanism of memory and attention (Bordalo, Gennaioli, and Shleifer, 2020), whereby the physician forms a norm based on similar and recent experiences and adjusts according to the difference between the norm and the

[40]Because of data availability, we mainly rely on results based on the ED setting to examine alternative explanations for positive autocorrelation discussed in the literature.

current choice situation.

The analysis in this paper suggests that physicians often rely on shortcuts, such as decisions about previous patients, to make fast decisions about current patients, especially when the consecutive patients are similar. While these shortcuts can help physicians make quick decisions given the time constraint, they also make them more vulnerable to biases that can have negative consequences for patients (Croskerry, 2002). In support of this view, we show that relying on the same decisions as for the previous patient shortens the consultation time for the current patient but increases unexpected discharge for high-risk patients.

Although it has been suggested that people use heuristics to make difficult decisions (Simon, 1955; Tversky and Kahneman, 1974; Gigerenzer and Todd, 1999), there has been a lack of welfare models to analyze the cost and benefit of using heuristics. On the benefit side, heuristics can be a valuable tool for simplifying complex problems, providing an approximation of optimal decisions, and saving time and cognitive resources. This can be especially important in a medical setting where timely decisions are necessary. Consider the mechanism of memory and attention (Bordalo, Gennaioli, and Shleifer, 2020) in this setting. Similarity in patient characteristics provides useful clues for assessing the risk of the current patient, and it is also less costly to retrieve more recent and similar patients. Instead of assessing each patient separately, physicians adjust their perceived risk for the current patient based on retrieved experiences from previous patients.

However, heuristics can also have negative implications. While more recent and similar patients may be easier and faster for the physician to retrieve, their treatments may not provide sufficient information to assess the risk of the current patient. Even if physicians can retrieve sufficiently similar patients, they may not be able to properly adjust the assessed risk. Thus, the use of shortcuts enables physicians to make faster decisions with less time and cognitive cost, but it may inevitably lead to systematic biases and errors, as Tversky and Kahneman (1974) noted: "these heuristics are quite useful, but sometimes they lead to severe and systematic errors." Although presenting a formal micro-foundation is beyond the scope of our paper, we believe that our study provides some empirical regularities for future theoretical work.

While completely removing shortcuts from clinical judgments would be unrealistic and perhaps unnecessary (Graber, Gordon, and Franklin, 2002), it is essential to understand the use of shortcuts in relation to the trade-off between decision time and decision quality. The welfare implications of using shortcuts depend on how they are used, in what context, and whether they are used in combination with other decision-making tools. Researchers have begun to examine this issue. For example, medical education can increase physicians' understanding of the pros and cons of using shortcuts (Graber, 2009), and clinical decision support systems, e.g., artificial intelligence, can suggest diagnoses for consideration (Sloane and Silva, 2020). Therefore, we highlight the need to understand heuristics and biases and to examine various methods for reducing medical errors toward better physician decisions.

8 Data Availability Statements

The data and code underlying this research is available on Zenodo at https://doi.org/10.5281/zenodo.8242347.

References

Akaishi, Rei, Kazumasa Umeda, Asako Nagase, and Katsuyuki Sakai. 2014. "Autonomous Mechanism of Internal Choice Estimate Underlies Decision Inertia." *Neuron* 81 (1):195–206.

Anderson, Theodore Wilbur and Cheng Hsiao. 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics* 18 (1):47–82.

Baskett, Thomas F., Victoria M. Allen, Colleen M. O'Connell, and Alexander C. Allen. 2007. "Fetal Trauma in Term Pregnancy." *American Journal of Obstetrics and Gynecology* 197 (5):499–e1.

Betrán, Ana P., Maria R. Torloni, Jia-Jia Zhang, A.M. Gülmezoglu et al. 2016. "WHO Statement on Caesarean Section Rates." *BJOG: An International Journal of Obstetrics & Gynaecology* 123 (5):667.

Bhargava, Saurabh and Ray Fisman. 2014. "Contrast Effects in Sequential Decisions: Evidence from Speed Dating." *Review of Economics and Statistics* 96 (3):444–457.

Bhattacharya, Jay, Timothy Hyde, and Peter Tu. 2018. *Health Economics*. Bloomsbury Publishing.

Bindler, Anna and Randi Hjalmarsson. 2019. "Path Dependency in Jury Decision Making." *Journal of the European Economic Association* 17 (6):1971–2017.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience Theory of Choice under Risk." *Quarterly Journal of Economics* 127 (3):1243–1285.

———. 2013. "Salience and Consumer Choice." *Journal of Political Economy* 121 (5):803–843.

———. 2020. "Memory, Attention, and Choice." *Quarterly Journal of Economics* 135 (3):1399–1442.

———. 2022. "Salience." *Annual Review of Economics* 14 (1):521–544.

Brachet, Tanguy, Guy David, and Andrea M Drechsler. 2012. "The Effect of Shift Structure on Performance." *American Economic Journal: Applied Economics* 4 (2):219–46.

Card, David, Alessandra Fenizia, and David Silver. 2023. "The Health Impacts of Hospital Delivery Practices." *American Economic Journal: Economic Policy* 15 (2):42–81.

Caughey, Aaron B., Alison G. Cahill, Jeanne-Marie Guise, and Dwight J. Rouse. 2014. "Safe Prevention of the Primary Cesarean Delivery." *American Journal of Obstetrics and Gynecology* 210 (3):179–193.

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

———. 2018. “The Efficiency of Slacking Off: Evidence from the Emergency Department.” *Econometrica* 86 (3):997–1030.

Chandra, Amitabh, David Cutler, and Zirui Song. 2011. “Who Ordered That? The Economics of Treatment Choices in Medical Care.” In *Handbook of Health Economics*, vol. 2. 397–432.

Chandra, Amitabh, Benjamin Handel, and Joshua Schwartzstein. 2019. “Behavioral Economics and Health-Care Markets.” In *Handbook of Behavioral Economics: Applications and Foundations*, vol. 2. 459–502.

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

Charness, Gary and Dan Levin. 2005. “When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect.” *American Economic Review* 95 (4):1300–1309.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. “Decision Making Under the Gambler’s Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires.” *Quarterly Journal of Economics* 131 (3):1181–1242.

Chen, Yiqun. 2021. “Team-Specific Human Capital and Team Performance: Evidence from Doctors.” *American Economic Review* 111 (12):3923–3962.

Croskerry, Pat. 2002. “Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias.” *Academic Emergency Medicine* 9 (11):1184–1204.

Currie, Janet M., Jonathan Gruber, and Michael Fischer. 1995. “Physician Payments and Infant Mortality: Evidence from Medicaid Fee Policy.” *American Economic Review* 85 (2):106–111.

Currie, Janet M. and W. Bentley MacLeod. 2008. “First Do No Harm? Tort Reform and Birth Outcomes.” *Quarterly Journal of Economics* 123 (2):795–830.

———. 2017. “Diagnosing Expertise: Human Capital, Decision Making, and Performance Among Physicians.” *Journal of Labor Economics* 35 (1):1–43.

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

Damisch, L., T. Mussweiler, and H. Plessner. 2006. “Olympic Medals as Fruits of Comparison? Assimilation and Contrast in Sequential Performance Judgments.” *Journal of Experimental Psychology Applied* 12 (3):166–178.

Davey, Mary-Ann and James King. 2016. "Caesarean Section Following Induction of Labour in Uncomplicated First Births—A Population-based Cross-sectional Analysis of 42,950 Births." *BMC Pregnancy Childbirth* 16 (1):1–9.

Erev, Ido and Ernan Haruvy. 2013. "Learning and the Economics of Small Decisions." In *Handbook of Experimental Economics*, vol. 2. 638–700.

Fang, Hanming and Qing Gong. 2017. "Detecting Potential Overbilling in Medicare Reimbursement via Hours Worked." *American Economic Review* 107 (2):562–91.

Fenner, Dee E., Becky Genberg, Pavna Brahma, Lorri Marek, and John O.L. DeLancey. 2003. "Fecal and Urinary Incontinence after Vaginal Delivery with Anal Sphincter Disruption in an Obstetrics Unit in the United States." *American Journal of Obstetrics and Gynecology* 189:1543–1549.

Gigerenzer, Gerd and Peter M Todd. 1999. "Fast and Frugal Heuristics: The Adaptive Toolbox." In *Simple Heuristics That Make Us Smart*. Oxford University Press, 3–34.

Gong, Qing. 2017. "Physician Learning and Treatment Choices: Evidence from Brain Aneurysms." *Working Paper* .

Graber, Mark. 2009. "Educational Strategies to Reduce Diagnostic Error: Can You Teach This Stuff?" *Advances in Health Sciences Education* 14 (1):63–69.

Graber, Mark, Ruthanna Gordon, and Nancy Franklin. 2002. "Reducing Diagnostic Errors in Medicine: What's the Goal?" *Academic Medicine* 77 (10):981–992.

Groopman, Jerome. 2007. *How Doctors Think*. Houghton Mifflin.

Herr, Paul M., Steven J. Sherman, and Russell H. Fazio. 1983. "On the Consequences of Priming: Assimilation and Contrast Effects." *Journal of Experimental Social Psychology* 19 (4):323–340.

Hohwy, Jakob. 2013. *The Predictive Mind*. Oxford University Press.

Kahana, Michael Jacob. 2012. *Foundations of Human Memory*. Oxford: Oxford University Press.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kahneman, Daniel and Dale T Miller. 1986. "Norm Theory: Comparing Reality to Its Alternatives." *Psychological Review* 93 (2):136–153.

Li, Jing, William H Dow, and Shachar Kariv. 2017. "Social Preferences of Future Physicians." *Proceedings of the National Academy of Sciences* 114 (48):E10291–E10300.

Litinski, Mikhail, Frank A.J.L. Scheer, and Steven A. Shea. 2009. "Influence of the Circadian System on Disease Severity." *Sleep Medicine Clinics* 4 (2):143–163.

Mullainathan, Sendhil and Ziad Obermeyer. 2022. "Diagnosing Physician Error: A machine learning Approach to Low-value Health Care." *Quarterly Journal of Economics* 137 (2):679–727.

New York State Department of Health (NYSDOH). 2015. "New York Statewide Planning and Research Cooperative System (SPARCS) Dataset, 2005–2015." Available at: https://www.health.ny.gov/statistics/sparcs/access/. Accessed June 1, 2020.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49 (6):1417–1426.

Radbruch, Jonas and Amelie Schiprowski. 2022. "Interview Sequences and the Formation of Subjective Assessments." *Working Paper* .

Sabbatini, Amber K., Brahmajee K. Nallamothu, and Keith E. Kocher. 2014. "Reducing Variation in Hospital Admissions from the Emergency Department for Low-Mortality Conditions May Produce Savings." *Health Affairs* 33 (9):1655–1663.

Shi, Pengyi, Mabel C Chou, JG Dai, Ding Ding, and Joe Sim. 2015. "Models and Insights for Hospital Inpatient Operations: Time-Dependent ED Boarding Time." *Management Science* 62 (1):1–28.

Silver, David. 2021. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *Review of Economic Studies* 88 (3):1385–1417.

Simon, Herbert. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1):99–118.

Simonsohn, Uri. 2006. "New Yorkers Commute More Everywhere: Contrast Effects in the Field." *Review of Economics and Statistics* 88 (1):1–9.

Simonsohn, Uri and George Loewenstein. 2006. "Mistake# 37: The Effect of Previously Encountered Prices on Current Housing Demand." *Economic Journal* 116 (508):175–199.

Singh, Manasvini. 2021. "Heuristics in the Delivery Room." *Science* 374 (6565):324–329.

Sloane, Elliot B and Ricardo J Silva. 2020. "Artificial Intelligence in Medical Devices and Clinical Decision Support Systems." In *Clinical Engineering Handbook (Second Edition)*. Academic Press, 556–568.

Southeast Asian Anonymous Hospital (SAAH). 2013. "Emergency Department Administrative Data, 2011–2012." Unpublished Data. Accessed January 29, 2018.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185 (4157):1124–1131.

Zhao, Haiyan, Björn Andersson, Boliang Guo, and Tao Xin. 2017. "Sequential Effects in Essay Ratings: Evidence of Assimilation Effects Using Cross-Classified Models." *Frontiers in Psychology* 8:933.

Figures and Tables

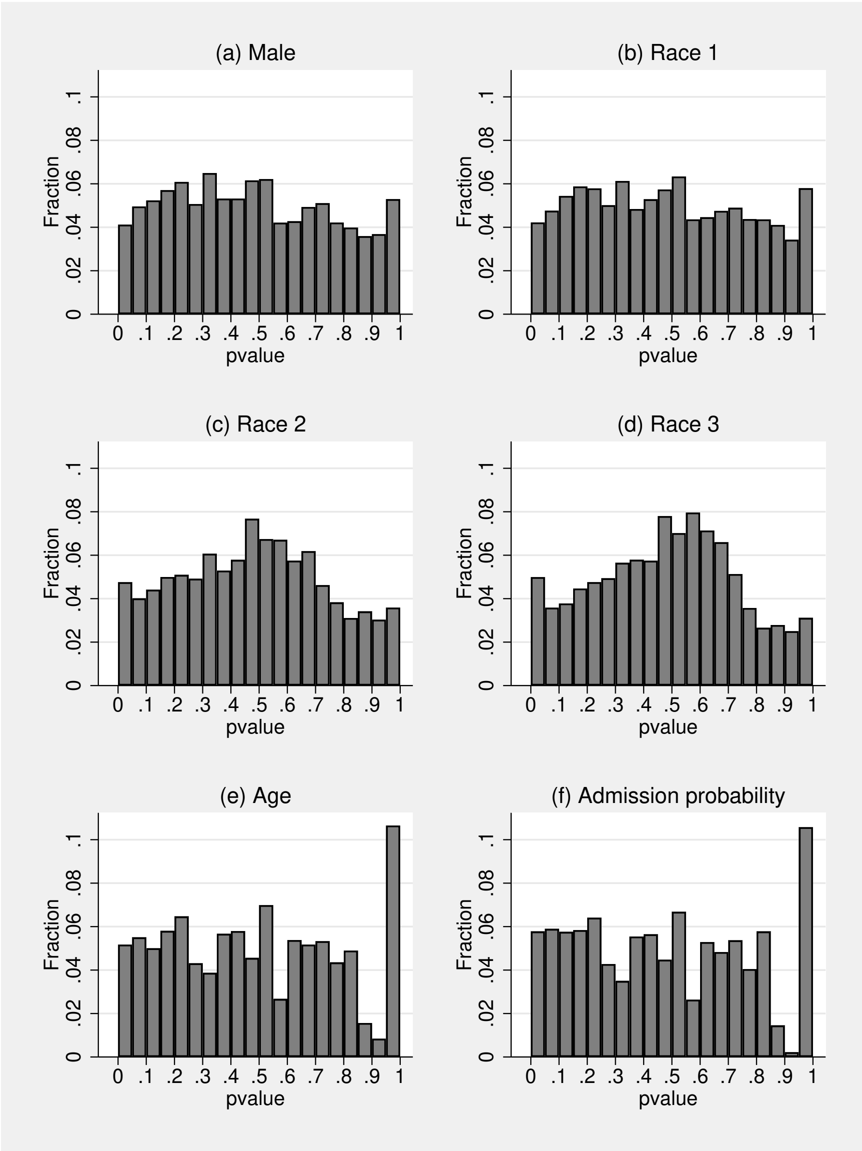

FIGURE 1

Distribution of P-values from Nonparametric Runs Tests of Patient Characteristics

Note: This figure presents the distribution of the runs test p-values for each physician-shift in the ED. Panels (a)–(f) consider the null hypotheses that the patient sequence follows a random order with respect to (a) patient gender, (b) Race 1, (c) Race 2, (d) Race 3, (e) patient age, and (f) admission probability predicted by patient demographics, triage severity, and diagnostic categories.

TABLE 1 Summary Statistics

Variable	Observations	Mean	SD
Panel A: Physician decisions[a]			
Inpatient admission	253,466	0.217	0.412
Order of lab tests	253,466	0.519	0.500
Order of imaging tests	253,466	0.512	0.500
Use of opioids	253,466	0.080	0.271
Use of antibiotics	253,466	0.052	0.221
Panel B: Patient characteristics[b]			
Male	253,466	0.647	0.478
Age	253,466	39.343	20.549
Race group			
Race 1	253,466	0.545	0.498
Race 2	253,466	0.199	0.399
Race 3	253,466	0.162	0.369
Others	253,466	0.094	0.292
Triage severity level			
1	253,466	0.038	0.191
2	253,466	0.247	0.431
3	253,466	0.715	0.451

Note: This table summarizes physician decisions and patient characteristics in the ED setting.

[a] Inpatient admission, order of lab tests, order of imaging tests, use of opioids, and use of antibiotics are dummy variables that measure, respectively, whether the physician orders inpatient admission, lab tests, imaging tests, opioids, and antibiotics for the patient.

[b] Unlisted patient characteristics include patient diagnostic categories.

TABLE 2 Autocorrelation in Disposition Decisions

	(1)	(2)	(3)
Y	Inpatient admission dummy		
Sample	Full sample	Same-shift patients	First patient in a shift
Panel A. Fixed-effect estimates			
Lag admission	0.0320***	0.0375***	0.0012
	(0.0026)	(0.0030)	(0.0070)
R-squared	0.398	0.392	0.460
Panel B. Pooled OLS estimates (no physician fixed effects)			
Lag admission	0.0383***	0.0438***	0.0083
	(0.0030)	(0.0034)	(0.0070)
R-squared	0.394	0.387	0.449
Patient demographics	YES	YES	YES
Triage severity	YES	YES	YES
Diagnosis	YES	YES	YES
Time fixed effects	YES	YES	YES
Observations	253,466	241,191	12,275
Sample mean outcome	0.217	0.207	0.414

Note: This table examines the effect of the physician's previous disposition decision on her current disposition decision in the ED. Panel A reports fixed-effect estimates from Equation (1). Panel B reports pooled OLS estimates without controlling for physician fixed effects. The dependent variable is a dummy that equals one if the physician admitted the current patient for inpatient care and zero otherwise. Lag admission is an indicator for whether the physician made an inpatient-admission decision for the previous patient. Observations are at the physician $\times$ patient visit level. In the full sample, observations are restricted to patient visits managed by the same physician within 2 days. Column (1) shows results using the full sample. Column (2) restricts the sample to patients who follow another patient within the same shift by excluding the first patient in each shift. Column (3) includes only the first patient in each shift. We note that the mean admission rate in column (3) is higher than that in column (2). This is because severe patients have longer consultations than non-severe patients; a physician can treat more patients when her shift is in the urgent care section with non-severe patients, compared with when her shift is in the acute care area with severe patients. Thus, the share of severe patients of the first patients in a shift is larger than the share of severe patients of total patients. All regressions control for characteristics of the current patient visit, including patient demographic characteristics (gender, race, and age), triage severity levels, diagnostic categories, and time fixed effects (hour of day, day of week, and month by year). Standard errors in parentheses are clustered at physician level. *** indicates significance at the 1% level.

TABLE 3 Autocorrelation in Other Physician Decisions

Y	(1) Lab test	(2) Imaging test	(3) Opioid use	(4) Antibiotic use
Lag Y	0.0226***	0.0301***	0.0182***	0.0197***
	(0.0020)	(0.0034)	(0.0030)	(0.0036)
Patient demographics	YES	YES	YES	YES
Triage severity	YES	YES	YES	YES
Diagnosis	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
Physician fixed effects	YES	YES	YES	YES
Observations	241,191	241,191	241,191	241,191
R-squared	0.404	0.272	0.092	0.175
Sample mean outcome	0.512	0.510	0.079	0.048

Note: This table examines autocorrelation in other physician decisions in the ED. The regression specification is the same as Equation (1), but with different measures of physician decisions. Columns (1)–(4) examine autocorrelation in the ordering of lab tests, imaging tests, opioids, and antibiotics, respectively. Observations are restricted to patients who follow another patient within the same shift. All regressions control for characteristics of the current patient, physician fixed effects, and time fixed effects. Standard errors in parentheses are clustered at physician level. *** indicates significance at the 1% level.

TABLE 4 Heterogeneity Analysis: Clinical Uncertainty

	(1)	(2)	(3)	(4)
Y	Inpatient admission dummy			
Lag admission	0.0397***	0.0639***	0.1316***	0.0370***
	(0.0064)	(0.0054)	(0.0205)	(0.0032)
Lag admission ×	-0.0171*			
Low-variation condition	(0.0086)			
Lag admission ×		-0.0433***		
Low admission risk		(0.0058)		
Lag admission ×		-0.0403***		
High admission risk		(0.0061)		
Lag admission ×			-0.0242***	
(log) Length of consultation			(0.0047)	
Lag admission ×				-0.0264***
Advanced diagnostic imaging				(0.0098)
Patient demographics	YES	YES	YES	YES
Triage severity	YES	YES	YES	YES
Diagnosis	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
Physician fixed effects	YES	YES	YES	YES
Observations	32,642	241,191	241,191	241,191
R-squared	0.448	0.398	0.393	0.416
Sample mean outcome	0.527	0.207	0.207	0.207

Note: This table examines the association between autocorrelation in admission decisions and clinical uncertainty associated with the current patient. In column (1), we study a sample of patients in which the patient is diagnosed with one of the 15 most common conditions for inpatient admission in the US (Sabbatini, Nallamothu, and Kocher, 2014). *Low-variation condition* is a dummy variable indicating that the patient's condition is among the seven conditions with variations less than median level among the 15 conditions. Column (1) includes the *Low-variation condition* indicator and its interaction with the lagged admission. In column (2), we predict admission probability for each patient using a logistic model, with admission decision as the dependent variable and patient characteristics, physician fixed effects, and time fixed effects as explanatory variables. *Low admission risk* (*High admission risk*) is a dummy variable indicating that the patient's admission probability is less than 0.25 (more than 0.75). Column (2) includes the *Low admission risk* and *High admission risk* indicators and their respective interaction with the lagged admission. Column (3) includes length of consultation in logarithmic form and its interaction term with the lagged admission. Column (4) includes the indicator for the use of advanced diagnostic imaging and its interaction term with the lagged admission. Observations are restricted to patients who follow another patient within the same shift. All regressions control for characteristics of the current patient, physician fixed effects, and time fixed effects. Standard errors in parentheses are clustered at physician level. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

TABLE 5 Heterogeneity Analysis: Physician Experience and Fatigue

	(1)	(2)	(3)
Y	Inpatient admission dummy		
Lag admission	0.0449***	0.0318***	0.0579***
	(0.0041)	(0.0039)	(0.0104)
Lag admission	-0.0140*		
× Experienced physician	(0.0079)		
Lag admission		0.0014**	
× Hours worked		(0.0006)	
Lag admission			-0.0061**
× (log) Shift time gap			(0.0029)
Patient demographics	YES	YES	YES
Triage severity	YES	YES	YES
Diagnosis	YES	YES	YES
Time fixed effects	YES	YES	YES
Physician fixed effects	YES	YES	YES
Observations	241,191	241,191	239,970
R-squared	0.392	0.392	0.392
Sample mean outcome	0.207	0.207	0.207

Note: This table examines whether experience and fatigue affect the degree of autocorrelation in physicians' decisions. Column (1) adds an indicator for experienced physician and its interaction with the lagged admission decision to Equation (1). Experienced physician indicates whether the physician's experience is at least 7 years. Column (2) includes the number of hours the physician has worked before treating the current patient in the shift and its interaction with the lagged admission decision. Column (3) includes the length of rest period before starting the current shift in logarithmic form and its interaction with the lagged admission decision. Observations are restricted to patients who follow another patient within the same shift. Dependent variables and other control variables are the same as those in Panel A of Table 2. Standard errors in parentheses are clustered at the physician level. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

TABLE 6 Effect of Path Dependency on Length of Consultation

	(1)	(2)	(3)
Y	(log) Length of consultation		
Repeat	-0.1322***	-0.1297***	-0.1498***
	(0.0095)	(0.0082)	(0.0102)
Patient demographics	YES	YES	YES
Triage severity	YES	YES	YES
Diagnosis	YES	YES	YES
Time fixed effects	YES	YES	YES
Physician fixed effects	NO	YES	YES
Admission	NO	NO	YES
Observations	241,191	241,191	241,191
R-squared	0.270	0.326	0.326
Sample mean outcome	3.579	3.579	3.579

Note: This table examines the effect of path dependency on the length of consultation. *Repeat* is a dummy that equals 1 when the physician made the same disposition decision for the current and previous patient and 0 otherwise. The dependent variable is the current patient's length of consultation in logarithmic form. Observations are restricted to patients who follow another patient within the same shift. Column (1) controls for characteristics of the current patient and time fixed effects. Column (2) further introduces controls for physician fixed effects and column (3) additionally controls for the current patient's disposition. Standard errors in parentheses are clustered at physician level. *** indicates significance at the 1% level.

TABLE 7 Autocorrelation and Unexpected Dispositions

Y	(1) 15-day admission	(2) 15-day ambulance	(3) 15-day admission	(4) 15-day ambulance	(5) Unexpected admission	(6) Unexpected discharge
Unexpected admission	-0.0005	-0.0002				
	(0.0012)	(0.0008)				
Unexpected discharge			0.0147***	0.0050***		
			(0.0018)	(0.0015)		
Lag admission					0.0133***	-0.0152***
					(0.0021)	(0.0023)
Patient demographics	YES	YES	YES	YES	YES	YES
Triage severity	YES	YES	YES	YES	YES	YES
Diagnosis	YES	YES	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES	YES	YES
Physician fixed effects	YES	YES	YES	YES	YES	YES
Observations	241,191	241,191	241,191	241,191	241,191	241,191
R-squared	0.019	0.024	0.019	0.024	0.034	0.194
Sample mean outcome	0.023	0.012	0.023	0.012	0.071	0.073

Note: This table examines whether autocorrelation in sequential physician decisions affects unexpected dispositions. *Unexpected admission* indicates whether the physician admits the current patient who is expected to be discharged. *Unexpected discharge* indicates whether the physician discharges the current patient who is expected to be admitted. *15-day admission* indicates whether the patient is admitted to hospital through the ED within 15 days of the current visit. *15-day ambulance* indicates whether the patient uses an ambulance to the ED within 15 days of the current visit. Observations are restricted to patients who follow another patient within the same shift. All regressions control for characteristics of the current patient, physician fixed effects, and time fixed effects. Standard errors in parentheses are clustered at physician level. *** indicates significance at the 1% level.

TABLE 8 Autocorrelation in Delivery Decisions

	(1)	(2)	(3)	(4)
Y	C-section dummy			
Sample	Full sample	Previous delivery was on same day	Weekend deliveries	Excluding likely-scheduled C-sections
Panel A. Fixed-effect estimates				
Lag C-section	0.0134***	0.0276***	0.0145***	0.0137***
	(0.0006)	(0.0011)	(0.0013)	(0.0007)
R-squared	0.334	0.404	0.217	0.066
Panel B. Pooled OLS estimates (no physician fixed effects)				
Lag C-section	0.0435***	0.0505***	0.0436***	0.0439***
	(0.0014)	(0.0016)	(0.0017)	(0.0014)
R-squared	0.336	0.405	0.219	0.068
Patient demographics	YES	YES	YES	YES
Patient conditions	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
Observations	2,458,773	828,034	524,781	1,985,701
Sample mean outcome	0.338	0.314	0.256	0.211

Note: This table examines the effect of the physician's previous delivery decision on her current delivery decision for childbirth. Panel A reports fixed-effect estimates from Equation (1). Panel B reports pooled OLS estimates without controlling for physician fixed effects. The dependent variable is a dummy that equals one if the physician performed a C-section on the current patient and zero otherwise. *Lag C-section* is an indicator for whether the physician performed a C-section on the previous patient. Observations are at physician $\times$ delivery level. Column (1) shows results using the full sample. Column (2) restricts the sample to deliveries in which the physician's previous delivery was on the same day as the current delivery. Column (3) restricts the sample to weekends, in which almost all deliveries are unscheduled emergency deliveries. Column (4) excludes patients with either a previous C-section or breech birth, which are two strong predictors of scheduled C-sections. All regressions control for characteristics of the current delivery, including patient demographic characteristics (race and age), 11 medical conditions listed in Appendix Table B.3, and time fixed effects (hour of day, day of week, and month by year). Standard errors in parentheses are clustered at physician level. *** indicates significance at the 1% level.

TABLE 9 Patient Similarity and Decision Autocorrelation

	(1)	(2)
Y	Inpatient admission dummy	
β_1: Lag admission × Group 1	0.0119	0.0055
[Nonsimilar severity & Nonsimilar disease & Nonsimilar age]	(0.0085)	(0.0089)
β_2: Lag admission × Group 2	0.0489**	0.0418*
[Nonsimilar severity & Nonsimilar disease & Similar age]	(0.0211)	(0.0216)
β_3: Lag admission × Group 3	0.0448	0.0369
[Nonsimilar severity & Similar disease & Nonsimilar age]	(0.0281)	(0.0293)
β_4: Lag admission × Group 4	-0.0757*	-0.0791*
[Nonsimilar severity & Similar disease & Similar age]	(0.0405)	(0.0409)
β_5: Lag admission × Group 5	0.0420***	0.0358***
[Similar severity & Nonsimilar disease & Nonsimilar age]	(0.0039)	(0.0036)
β_6: Lag admission × Group 6	0.0545***	0.0482***
[Similar severity & Nonsimilar disease & Similar age]	(0.0058)	(0.0055)
β_7: Lag admission × Group 7	0.0538***	0.0477***
[Similar severity & Similar disease & Nonsimilar age]	(0.0073)	(0.0068)
β_8: Lag admission × Group 8	0.1036***	0.0980***
[Similar severity & Similar disease & Similar age]	(0.0127)	(0.0126)
Group FE	YES	YES
Patient demographics	YES	YES
Triage severity	YES	YES
Diagnosis	YES	YES
Time fixed effects	YES	YES
Physician fixed effects	YES	YES
Observations	241,191	241,191
R-squared	0.388	0.392
Sample mean outcome	0.207	0.207

Note: This table reports estimation results from Equation (5). Consecutive patients are divided into 8 groups based on their similarity in three dimensions—similar severity or not × similar disease or not × similar age or not. The explanation in square brackets under "*Lag admission × Group #*" indicates the similarity status in the corresponding group. For example, *Group 1* indicates that the current and previous patient are not similar in any of the three characteristics; the coefficient on "*Lag admission × Group 1*" measures the degree of autocorrelation in admission decisions for patients in group 1. Observations are restricted to patients who follow another patient within the same shift. Column (1) controls for characteristics of the current patient, time fixed effects, and group fixed effects. Column (2) additionally controls for physician fixed effects. Standard errors in parentheses are clustered at physician level. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

TABLE 10 Parameters from Structural Estimation

Parameter	Value	SE
λ_{SEV}	3.4790	0.1183
λ_{DZ}	1.4977	0.1572
λ_{AGE}	1.6767	0.1065
α	0.0098	0.0003
σ	64.7541	0.6766

Note: This table presents estimates of the parameters in Equation (10) using maximum likelihood estimation. Standard errors are bootstrapped.