Decision Fatigue in Physicians

*B*y HAN YE, JUNJIAN YI, AND SONGFA ZHONG*

We explore the consequences of physician's excessive workload from the perspective of decision fatigue—the decline in decision quality due to an increased number of patients and decisions. Leveraging an administrative dataset of 240,000 emergency department visits, we find that increased number of patients decreases inpatient admission rates, task orders, and patient length of stay. Subsequently, both patient revisit rates and mortality rates increase. The results are robust if we use the number of ambulance arrivals as an instrumental variable. Furthermore, we find that the observed consequences in physician decision-making can be alleviated by taking a break and by accumulated medical experience. (JEL D91, I18, J44)

* Ye: Department of Economics, National University of Singapore, Singapore 117570 (email: yehan0413@gmail.com); Yi: Department of Economics, National University of Singapore, Singapore 117570 (email: junjian.yi@gmail.com); Zhong: Department of Economics, National University of Singapore, Singapore 117570 (email: zhongsongfa@gmail.com).

I. Introduction

Excessive workload of physicians has been a long-standing problem in medical care (Blendon et al., 2002; Dzau et al., 2018). This problem has worsened in recent years, largely driven by public pressure to reduce operating costs and an aging society that generates greater healthcare needs. Physicians in the United States, for example, on average treat 21 patients per day and work 53 hours per week (Physicians Foundation, 2016). Numerous studies have shown that a heavy workload impairs physicians' decision-making (Michtalik et al., 2013). In particular, a large portion of medical errors, estimated to cause more than 250,000 deaths each year in the United States, has been attributed to physicians' excessive workload (Makary and Daniel, 2016). Rising awareness of the workload problem has led to public debates regarding restrictions on residents' duty hours in many countries (Philibert et al., 2002; Nasca et al., 2010). While duty-hour restrictions reduce excessive workloads, policy makers are concerned that shorter shifts can potentially cause more inpatient handoffs and work compression (Gee, 2011).

This study examines the relationship between physician workload, decision-making, and the quality of administered medical care from a psychological perspective. Excessive workload for physicians is mostly characterized by the increased number of patients to be treated, and consequently the increased number of decisions to be made within a shift. One important consequence of excessive workload is decision fatigue, a notion recently proposed by behavioral scientists. More specifically, because decision-making requires mental resources that are in limited supply, decision quality declines after making a sufficiently long series of decisions (Baumeister et al. 1998; Vohs et al., 2008; Johnson, 2008). For example, the decision quality of consumers, financial analysts, voters, and even judges is negatively affected by the number of decisions they have previously made (Levav et al., 2010; Hirshleifer et al., forthcoming 2018; Augenblick and Nicholson, 2015; Danziger et al., 2011). Studies have also shown that students are more likely to underperform on standardized tests (Sievertsen et al., 2016), and clinicians prescribe unnecessary antibiotics (Linder et al., 2014) for every hour later in the day because of their taxed mental resources. Moreover, as anecdotal evidence suggests, Barack Obama, Steve Jobs, and Mark Zuckerberg, among others, have often adopted minimalist fashion to reduce decision fatigue.[1] In healthcare, an excessive workload compels physicians to make too many medical decisions, which can erode their decision quality and cause often dire consequences for patients. Hence understanding its psychological underpinnings is crucial to reduce overall medical and financial burdens.

We hypothesize that decision fatigue, indexed by the number of patients a physician treats in a given shift, potentially erodes physician decision quality and patient health outcomes. To test this hypothesis, we employ administrative data from a large emergency department (ED) in Singapore. Our dataset contains 242,761 patient visits with 128 physicians over a period of two years. This dataset uniquely fits our research objective in four respects. First, the hospital information system documents comprehensive records on all ED visits, including patient characteristics, physician decisions, patient outcomes, and, importantly, timestamps for the patient's path through the ED. Second, financial incentives are unlikely to play a role in the ED we studied, as physicians are paid a monthly salary with a fixed shift allowance, and patients incur a fixed attendance fee upon registration. Third, physicians and patients are generally randomly matched, due to the numerous modalities of cases that present for emergency care combined with physicians' predetermined shift schedules (Chan, 2016, 2018). Last, the centralized

---

[1] See "The scientific reason why Barack Obama and Mark Zuckerberg wear the same outfit every day." https://www.businessinsider.com/barack-obama-mark-zuckerberg-wear-the-same-outfit-2015-4/?IR=T.

ambulance system in Singapore ensures that ambulance arrivals at the ED are exogenous to hospital conditions. This exceptional institution serves as a quasi-experiment to identify the causal effects of decision fatigue, since ambulance arrivals contribute to a substantial proportion of physicians' workload.

We find a decrease in the probability of inpatient admission, the number of task orders, and patient length of stay as the number of patients a physician treats in a given shift increases. Controlling for various confounding factors, the ordinary least squares (OLS) estimates suggest that every ten patients the physician has previously treated during a shift lower the index patient's probability of hospital admission by 11.7%, reduce the number of task orders by 12.3%, and shorten the length of stay by 19.1%. We further examine the extent to which we can identify the causal effect in two separate analysis. First, we show that these results are robust after we address the potential non-random patient-physician assignment by using the exogenous ambulance arrivals to the ED . Second, we find that the consequences of decision fatigue in physician decision-making could be alleviated by taking a break, which is consistent with previous research demonstrating the beneficial effects of breaks on individuals' performance (Danziger et al., 2011; Sievertsen et al., 2016).

We further investigate whether physicians' decision fatigue aggravates patient treatment outcomes. We find that decision fatigue significantly erodes the quality of patient treatment, leading to a higher likelihood of patient return visits as well as ED mortality. Specifically, every ten patients previously treated in the physician's shift increase the index patient's revisit rates and mortality rates by 3.6% and 12.7%, respectively.

Finally, we study whether physicians' professional experience mitigates the effects of decision fatigue. We do this by first estimating the decision-fatigue effect for each physician, and then regressing the individual-specific estimate on the physician's characteristics. Our results suggest that medical experience mitigates the effects of decision fatigue for young physicians, although the effect diminishes as the physician ages.

Our study contributes to the understanding of the consequences of physician workload in several ways. First, the literature on physician workload exclusively focuses on working hours (Blendon et al., 2002; Linder et al., 2014; Dzau et al., 2018). Our paper is the first to investigate decision fatigue, measured by the number of patients a physician treats in a given shift. Our results remain significant both economically and statistically if we further control for the number of hours a physician has worked for, suggesting for a net effect of decision fatigue independent of working hours. We thus provide a new perspective to understand the long-standing problem of physician over-workload in medical care. Over-workload means not only the number of hours worked, but also the number of patients treated, or the number of decisions made in a given time period. Second, unlike previous studies on physician workload, our data enable us to identify the causal effect of decision fatigue. We use an instrumental variable (IV) method to examine the causality of decision fatigue on physician decision-making, exploiting the exogeneity of ambulance arrivals at the ED. Third, with patient treatment outcomes available, our study is able to assess whether and how decision fatigue affects the quality of physician decisions. Lastly, the physician-level information allows us to examine how physician characteristics, in particular, professional experience affects physicians' responsiveness to decision fatigue.

Our study also contributes to the increasing literature on the use of behavioral approaches to understand medical decision making and overcome challenges in healthcare (Wakker, 2008; Cohen et al., 2016; Li et al., 2017). Specifically, overworked physicians are vulnerable to cognitive biases and making suboptimal decisions (Chandra et al., 2011; West et al., 2016). Our study evaluates the effect of physicians' excessive workload from the psychological perspective of decision fatigue, which has been shown to affect the decision quality of consumers, financial analysts, voters, and even judges (Levav et al., 2010; Hirshleifer et al., forthcoming 2018; Augenblick and Nicholson, 2015; Danziger et al., 2011).

The psychological mechanisms underpinning decision fatigue perhaps can be best understood within the framework of two systems of cognitive processes (Kahneman and Egan, 2011). In this scheme, decisions arise either from the fast and effortless System 1 or the slow and effortful System 2. As physicians suffer from increasingly depleted mental resources toward the end of the shift, they have to rely more and more on System 1 to make fast and effortless decisions. Hence, they are more likely to make poorer decisions that lead to worse patient outcomes. Relatedly, our results are also consistent with the recent literature on the behavioral consequences of scarcity (Shah et al., 2012; Mani et al., 2013; Shafir and Mullainathan, 2013). They argue that resource scarcity impedes cognitive function, which in turn may lead to suboptimal decisions. In our context, increased number of patients and decisions may generate a sense of scarcity in terms of cognitive resources leading to diminishing performance of the physicians.

Our study further contributes to a growing literature on health economics analyzing physician decision-making. It is well documented that physicians' performances are determined not only by their human capital (Currie and MacLeod, 2017), but also by their surrounding environments (Chandra and Staiger, 2007; Skinner, 2011; Chan, 2016), including extraneous factors unrelated to patients' health. For instance, financial and liability considerations may sway physicians to perform unnecessary procedures (Currie and MacLeod, 2008; Clemens and Gottlieb, 2014). A recent work of Fang and Gong (2017) finds that physicians' financial incentives also affect their decisions of Medicare claims. Our paper is closely related to the work of Chan (2018), which examines two behavioral distortions in ED physicians due to work schedule. First, physicians accept fewer patients near end of shift (EOS). Second, physicians shorten the duration of care and increase formal utilization on patients assigned near EOS. Our study contributes to the understanding of how physicians make sequential decisions over the course of a shift.

The rest of this paper is organized as follows. Section 2 describes the institutional background and discusses the dataset. In Section 3, we investigate the effect of decision fatigue on physician behavior. In Section 4, we study the effect of decision fatigue on patient treatment outcomes. In Section 5, we move to examine the role of professional experience in physicians' responsiveness to decision fatigue. Section 6 concludes.

II. Institutional Setting and Administrative Data

In this section, we describe institutional background, introduce patient flow and physician shifts, and define our main variables based on the administrative data.

A. Institutional Setting for Identification

A key challenge in identifying the effect of decision fatigue on physician decision-making is the endogenous matching between patients and physicians. In most healthcare settings, patients are not randomly assigned to physicians: Not only patients search for physicians, but also physicians choose their patients (Lu and Rui, 2017). By contrast, the ED in our research setting offers three distinct advantages to address this challenge.

Patients and physicians are almost randomly matched in the ED (Chan, 2016, 2018). The rationale is twofold. First, ED visits are unplanned. Patients are not likely to select their physicians due to the unexpected nature of emergency care, and they are assigned by a triage nurse to on-shift physicians upon arrival. Second, the internal shift scheduling of physicians is predetermined, and physicians cannot control the volume of ED arrivals or the types of patients assigned to them by the triage nurse. As a result, the match between patients and physicians is largely random.

In addition, ambulance system is centralized in our setting, which ensures that ambulance arrivals are exogenous to hospital conditions and patient characteristics. Singapore's emergency medical services system is operated by the Singapore Civil Defense Force (SCDF). As shown in Table A1, SCDF ambulances transport more than 93% of the ambulance arrivals in our data. SCDF ambulance personnel conveys patients only to the nearest hospital, and will not consider requests to redirect patients to alternative hospitals. As a result, the number of ambulance arrivals should be independent of hospital characteristics and patient conditions. However, patients who arrive by ambulance are a major determinant of ED physicians' workload. This unique institutional feature allows us to identify causal effects of decision fatigue, using the quasi-experimental variation in physicians' workloads induced by ambulance arrivals.

Finally, physicians' decisions in the ED we study are not influenced by financial incentives. Government subsidies are provided for every ED patient regardless of nationality, and all patients incur a fixed attendance fee upon registration. Physicians are paid a basic monthly salary with a fixed shift allowance, and are compensated by neither the quality nor the quantity of work within the scheduled shift.

B. Patient Flow and Physician Shifts

Patient Flow.—Figure A1 illustrates the patient flow process in the ED. Upon arrival at the ED, patients are registered, screened, and triaged by a triage nurse. Triage is based on a four-level patient acuity category scale (PACS), with level 1 being the most severe cases, level 2 major emergencies, level 3 minor emergencies, and level 4 non-emergency cases. A scheduling system then determines the assignment of patients to each on-shift physician and the order of their consultations on a first-come, first-served basis. Patients with severe conditions (PACS levels 1 and 2, henceforth severe cases) have higher priority than the rest. With few exceptions leaving after initial consultations, most patients undergo some type of diagnostic testing, such as lab work or X-rays, or receive treatment by a nurse or physician assistant. When test results are available, or the treatment is completed, the patient is reviewed by the same physician before being discharged or hospitalized.

The administrative dataset records real-time patient flow in the ED. It is organized by patient visits, with each record corresponding to a single visit. Each record contains detailed timestamps on the patient's complete path through the ED, such as when a patient arrived at the ED, when the patient was seen by a physician, when the physician ordered any test or treatment, and when the physician made a final discharge disposition. For each visit, the physician who carried out patient care is identified by a unique ID. Since the dataset records information on clinical workflow for all visits, we are able to track the universe of physicians' activities in the ED.

Physician Shifts.—Following the procedure in Brachet et al. (2012), we construct physicians' shifts based on their periods of inactivity, which is identified by their absence from the administrative data. Sorting the 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 six or more hours have elapsed between consecutive observations of the same physician. The rationale behind the six-hour cutoff between visits to define a new shift is as follows: First, it is almost impossible that a physician would be on duty for six consecutive hours without a single case, given overcrowding and long waits in the ED; second, a physician's rest period between two consecutive shifts is unlikely to be less than six hours.

The shifts we identified from the data may differ from actual shift schedules. For example, if on-duty physicians remain inactive for six or more hours, our procedure will assign them a new shift, although they could still be on the same long shift. However, this type of misclassification does not pose any threat to the validity of our estimates. The effects we aspire to identify arise from the decision fatigue that results from making repeated patient care decisions, rather than long on-duty hours. A physician may be well rested after long hours that did not include patient care, in which case it is reasonable to define a new shift for the purposes of our study. In another example, if the rest between two consecutive shifts is less than six hours, we classify the physician as being on a longer shift; we also use four-hour and five-hour cutoffs to define new shifts as robustness checks. The results presented below are not sensitive to these alternative definitions.

Once the shifts are defined, we measure the duration of a shift as the number of hours elapsed (rounded up to the nearest integer) from the start of the first patient's consultation to the end of the last consultation in the shift. Figure A2 plots the distribution of shift durations. The most frequently occurring mode is eight hours, and around half of the shifts are longer than eight hours. The actual hours worked may differ from the planned work schedules, as ED physicians may have unpredictable work schedules for unforeseen circumstances. For example, physicians are expected to work beyond scheduled shifts for reasons such as task completion and staffing shortages (Morrow et al., 2014).

The dataset contains 264,115 patient visits to the ED over two years—from January 1, 2011, to December 31, 2012. We construct physician shift schedules and real-time patient flow volume using all visits, and focus on a restricted sample for analysis. This sample includes patient visits for which (i) the physician has at least ten shifts observed during our sample period, and (ii) the physician is working in a shift with duration between 6 and 16 hours. By placing restrictions on physician shifts, we rule out the possibility of unstable temporary staffs and unusual working hours. We also exclude cases of death on arrival before being assigned to a physician. Our final sample contains 242,761 patient visits with treatments by 128 physicians.

C. Main Variables and Summary Statistics

Decision Fatigue.—During our study period, physicians averagely treat 21 patients per shift and work 42 hours per week in the ED. These physicians' workloads are comparable to those observed in the literature on excessive physician workload (Physicians Foundation, 2016). Different from the literature that measures physicians' over-workload by long working hours, we characterize excessive workload by the overwhelming number of decisions, such as task orders and treatment decisions. Previous studies in decision science and behavioural economics suggest that making sequential decisions depletes individuals' executive functioning and causes mental fatigue, which can influence subsequent decisions (Levav et al., 2010; Danziger et al., 2011; Augenblick and Nicholson, 2015; Hirshleifer et al., forthcoming 2018). We measure decision fatigue as the number of patients the physician treats in a given shift. This ordinal position serves as a proxy for the real-time cumulative workload and a measure of the physician's cognitive depletion. As shown in Panel A of Table 1, on average, a physician has treated 10.4 patients before the index patient's consultation in a given shift.

Physician Decisions.—We have three measures for physician decision-making: (i) inpatient admission, (ii) number of task orders, and (iii) patient length of stay. Panel B of Table 1 presents summary statistics for these three decision variables.

Physician's discharge decision is the primary product of ED care and a matter of discretion for physicians (Chan, 2018). After the completion of immediate resuscitation and treatment, the physician may discharge a patient home or refer to a primary care center for follow-ups. If a patient has a serious condition, the physician may admit the patient to a specialist department in the hospital. The admission disposition may not occur at the end of ED care, as inter-unit handoffs from the ED to inpatient care require coordination between different parties. Patients awaiting hospital admission may have to remain in the ED for at least several hours. We focus on the decision to admit the patient as a key outcome measure, which accounts for 17.6% of the sample visits.

We also examine physician input of care for task orders and consultation time. Medical task orders include treatments, procedures, tests, and medications. We count the total number of task orders to measure hospital resource utilization. We measure the length of stay as the minutes elapsed from the start to the end of the patient's consultation to a specific physician. This time duration includes the time for history-taking, initial examination, formal tasks, and review of test results, but excludes waiting times for initial consultation and admission to an inpatient ward. As shown in Panel B of Table 1, on average, a patient receives 5.3 tasks, and stays for approximately one hour in the ED.

Patient Outcomes.—We focus on two measures of patient outcomes: ED revisits and mortality. Specifically, ED revisits measure whether a patient revisited the ED within 14 days.[2] ED mortality indicates whether a patient died in the ED after assigned to a physician. Panel C of Table 1 reports summary statistics for these two variables. The 14-day revisit rate is 12.9% and the ED mortality rate is 0.2%. All ED deaths belong to those triaged as severe cases. Although these two measures largely depend on a variety of factors outside the ED, we seek to eliminate potential confounding factors by comprehensively controlling for patient characteristics, physician fixed effects, and time fixed effects.

[2] We identify multiple visits for the same patient using comprehensive patient information including gender, race, birth date and home address.

Patient Characteristics.—We observe much of the information available to the physician at the time of accepting the patient visit, including the patient's gender, race, age, and triaged severity level. Panel D of Table 1 reports summary statistics for these ex ante patient characteristics. In our data, 65% of the patients who visited the ED during the sample period are men. The average age of patients is around 39 years. More than 70% of patient visits are minor emergency cases, which are classified as PACS levels 3 (henceforth non-severe cases); the remaining 4% are level 1 cases, and 24% are level 2.

We also have ex post diagnostic information for each patient. Physicians make diagnostic judgments after interacting with patients or reviewing their test results. In some specifications, we include patient diagnostic category to predict clinical decisions. Diagnostic groupings and code numbers are based on the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM).

Patient arrivals at the ED are not smooth; we observe considerable fluctuations in ED occupancy over time. For example, Monday and Sunday are the busiest days within a week and 10 am to 3 pm and 8 pm to 11 pm are the two peak hours within a day. The total number of patient visits increased over the two years, and ED patient volumes varied across months. We include a set of time fixed effects in our regression analysis to account for time variations.

Physician Characteristics.—To test the relationship between experience and physicians' responsiveness to decision fatigue, we obtained a sample of 101 physicians out of the 128 physicians with physician characteristics including age, gender, education background, and, medical experience. Medical experience is defined as the number of years since a physician obtained his or her first degree to practice medicine. To capture training background, we include a dummy variable indicating whether a physician obtained the first medical degree from a local university or from an oversea university. Compared to graduates abroad, physicians trained at home may have better knowledge of local demographics and institutional regulations. We also measure whether the physician received continuing medical education after completing the initial medical training. Continuing medical education is defined as the acquisition of professional medical credentials after the initial graduation from medical school.

Panel E of Table 1 presents summary statistics for physician characteristics. On average, a physician has been in practice for six years since obtaining the basic medical degree. The majority of ED physicians are males (77.2%). Around half (51.5%) obtained their initial medical training locally, and one quarter (24.8%) obtained other medical credentials after their initial medical education.

III. Impacts of Decision Fatigue on Physician Behavior

In this section, we conduct regression analyses to quantify the effects of decision fatigue on physician behavior while controlling for patient demographics, case severity, physician fixed effects, and a series of time fixed effects. We also examine whether and how taking a break alleviates the consequences of decision fatigue. To address concerns regarding non-random patient assignment, we further explore a unique institutional feature to establish the causal link between decision fatigue and physician decisions. Finally, we assess the robustness of our results by performing an extensive array of sensitivity analyses, and explore the heterogeneous effects of decision fatigue on physician behavior.

A. Baseline Regressions

Panels A-C in Figure 1 plot physician decisions by the number of patients previously treated within the shift. In Panel A, we observe a substantial decline in the likelihood of hospital admission as the number of patients treated by the physician increases. Panels B and C show that physicians reduce task orders and shorten patient length of stay after they have treated more patients in a shift. To rigorously examine the graphical pattern, we conduct regression analyses controlling for multiple confounding factors.

Regression Specification.—We start from a linear model in which we assume that neither patients nor triage nurse select physicians in the ED; instead, patient visits are randomly assigned to each on-shift physician. The baseline regression that describes the association between decision fatigue and physician behavior is:

$$(1) \qquad Y_{ijt} = \alpha PatientCount_{ijt} + X_i\beta + T_t\gamma + \nu_j + \epsilon_{ijt},$$

where medical decision Y_{ijt} is indexed for patient visit i treated by physician j starting consultation at time t. For the admission decision, Y_{ijt} is a dummy variable that equals one if patient i is admitted to the hospital and zero otherwise. For the measure of task orders, Y_{ijt} represents the total number of task orders for patient i. For patient length of stay, Y_{ijt} takes log transformation. We fit linear models for all outcomes, and conduct an additional probit regression for the binary admission decision.

The object of interest is physician decision fatigue—$PatientCount_{ijt}$, which counts the number of patients seen by physician j during the shift before the start time t of patient i's consultation. X_i is a vector of patient demographic characteristics and severity levels. Demographic variables include patient gender, race, age, and age squared. Patient severity is indexed by PACS levels. We also control for time fixed effects T_t and physician identities ν_j. Time fixed effects include hour of day, day of week, and month-year interactions. The error term, ϵ_{ijt}, captures measurement errors. Because of potential serial correlations for patients treated by the same physician, we cluster standard errors at the physician level throughout.

Regression Results.—Panel A of Table 2 shows estimation results from Equation (1). All models estimate statistically significant and negative coefficients on decision fatigue.[3] The results suggest that the declining patterns in the left panels of Figure 1 hold even after controlling for patient case attributes, time categories, and physician fixed effects.

Columns (1) and (2) in Panel A of Table 2 present results for hospital admission. The ordinary least squares (OLS) estimate for α in Column (1) is -0.0021 (standard error, 0.0002). This estimate suggests that every ten patients the physician has previously treated during the shift lower the index patient's hospital admission probability by 11.7%, from a sample mean likelihood of 0.176. The average marginal effect from the probit estimation is -0.0029 (standard error, 0.0004), as presented in Column (2). There are no significant differences between OLS estimates and marginal effects from the probit model. Henceforth, we only discuss OLS estimates.

Columns (3) and (4) show estimates for decision-fatigue effects on task orders and length of stay. Physicians tend to reduce task orders and shorten patient consultation time as they advance in the sequence of patient cases. Specifically, every ten patients previously treated by the on-shift physician is associated with a 0.655-unit reduction of task orders, which is 12.3% less than a sample mean of 5.317 orders. Meanwhile, patient length of stay is shortened by 19.1% for every ten additional patients previously treated by the physician during the shift.

These results suggest that decision fatigue plays a significant role in physician decision-making. When they are mentally fatigued, physicians may tend to simplify medical decisions by lowering inpatient admissions, reducing task orders, and shortening the length of stay. Two factors support our view that inpatient admission

[3] For brevity, we only report the estimates of the key parameter, i.e. α. The unreported estimates of coefficients on other control variables have the expected sign and magnitude.

is a more complex decision than outpatient discharge, and thus a less likely outcome when decision fatigue increases. First, the physician must decide which specialist department is the most appropriate to admit the patient to for further treatment. This is not a straightforward decision, especially for patients with multiple medical problems. Second, the physician must coordinate with the specialist department for inter-unit handoffs (Apker et al., 2007). After approval of the admission request, the patient remains in the ED until an inpatient bed becomes available.

B. Breaks within a Shift

Psychological studies suggest that mental fatigue can be partially overcome by a short rest (Tyler and Burns, 2008; Danziger et al., 2011). Using the administrative data, we define a break as a period of at least one hour, during which the physician on shift is not in charge of any patient. The break divides a shift into distinct decision sessions. Here we restrict our analytic sample to patient visits with physicians who are working in a shift with one break. This restriction yields a working sample of 23,733 visits.

Panels D-F of Figure 1 plot physician decisions by the number of patients treated in each session, and demonstrate that the break restores the physician to a high level of functioning. As shown in Panel D, the likelihood of hospital admission steadily declines as the number of previously treated patients increases, but rebounds right after a break. Similarly, Panels E and F show that the number of task orders is larger, and the length of stay is longer at the very beginning of the shift or right after a break than later in the sequence of cases.

To rigorously examine the graphical pattern above, we extract three groups of patient visits from the analytic sample based on their positions relative to a break. Group 1 comprises the last fourth to sixth cases treated before the break, Group 2 the last three cases before the break, and Group 3 the first three cases after the break. Using these patient visits, we estimate the following equation:

(2) $$Y_{ijt} = \alpha_1 Group1_{ijt} + \alpha_2 Group3_{ijt} + X_i\beta + T_t\gamma + \nu_j + \epsilon_{ijt},$$

where $Group1_{ijt}$ and $Group3_{ijt}$ are two dummy variables, indicating whether patient i belongs to Group 1 or Group 3, respectively. The reference category refers to patients in Group 2.

Holding other factors constant, α_1 measures the difference in physician decisions between Group 1 and Group 2 patients, and α_2 is the difference in physician decisions between Group 3 and Group 2 patients. More specifically, $-\alpha_1$ reflects the effect on physician decisions when the number of previously treated patients increases by three in the same session, and α_2 represents the combined effect of treating three more patients and taking a break. Therefore, $\alpha_1 + \alpha_2$ captures the effect of a break.

Table 3 shows the estimation results from Equation (2). OLS estimates for α_1 and α_2, with only one exception, are statistically significant and positive. This result confirms the statisitcal robutness of the pattern depicted in the right panels of Figure 1, namely, the declining trend within a decision session and the restoration right after a break. Moreover, the effect of a break, $\alpha_1 + \alpha_2$, is estimated to be statistically significant and positive in all models. This result is consistent with the literature demonstrating that mental resources can be replenished by interventions such as a short rest (Danziger et al., 2011; Sievertsen et al., 2016).

C. Consideration of Non-random Work Assignment

A key concern for interpreting the association between physician decision fatigue and behavior embodied in Equation (1) as a causal relationship is that the assignment of patients to physicians might not be random. Although the ED provides a context in which patients and physicians are nearly randomly matched, the risk of causal inference in OLS estimates remains. For example, triage nurses may observe the degree of physicians' decision fatigue and assign fewer complicated cases to more fatigued physicians. If this were the case, our OLS estimates would be biased.

We address this concern by exploring a unique feature of Singapore's centralized emergency ambulance system. As described in the previous section, the number of ambulance arrivals strongly predicts the volume of work in the ED, but is orthogonal to hospital conditions and patient health. Therefore, we use the total number of ambulance arrivals at the ED during the physician's shift up to the arrival of the index patient as an IV for the number of patients previously treated by the physician. Panel F of Table 1 shows that averagely ten patients arrived at the ED by ambulance from the physician's shift start to the consultation of the index patient. First-stage regressions, shown in Panel B of Table 2, demonstrate a strong positive correlation between the number of hospital ambulance arrivals and a physician's workload.

Panel C of Table 2 shows that the signs of IV estimates are consistent with OLS estimates, but their magnitudes are substantially larger. Specifically, every ten patients previously treated cause the on-shift physician to lower the index patient's hospital admission probability by 19.4%, reduce the number of task orders by 19.1%, and shorten the length of stay by 23.0%. Moreover, the results of the Hausman test show that the differences between OLS and IV estimates are statistically significant at the 1% level. This result suggests that OLS estimates are biased downward, perhaps because triage nurses take into account physicians' decision fatigue when assigning patients to physicians. However, the effect of decision fatigue on physicians remains substantial, as the sizable OLS estimates suggest.

D. Robustness Analyses

Decision Fatigue vs. Physical Fatigue.—Our measure of decision fatigue—the number of patients previously treated in the shift—is correlated with elapsed time in the given shift. That is, physicians treat more patients as the shift wears on. One might thus be concerned that the observed patterns are actually due to working hours rather than decision fatigue; physicians could behave differently simply because of longer working hours. To address this concern, we conduct an analysis that includes cumulative hours elapsed in the physician's shift as an additional control variable.

Panel A of Table 4 reports regression results after controlling for cumulative time. Coefficients on the number of patients treated remain negative and statistically significant in all models, although the magnitudes are smaller than those in the baseline regressions. This result suggests that the behavioral differences exhibited by the physician are not simply due to elapsed time. Fixing the number of working hours, repeated decision-making still exhibits a statistically and economically significant effect on physicians' subsequent decisions. In particular, this interpretation should be viewed in light of the high correlation between the number of patients treated and cumulative hours (Pearson correlation=0.68, P<0.001).

End-of-Shift Effect.—Previous studies have provided evidence on performance deterioration near the end of workers' shifts (Brachet et al., 2012; Chan, 2018). For example, Chan (2018) finds that ED physicians order more formal tasks and complete their work earlier as the end of the shift approaches. To check whether our estimates of decision fatigue are driven by end-of-shift effects, we exclude the last three patient visits in each shift as a robustness analysis.

Panel B of Table 4 presents regression results for this restricted sample. Estimates remain almost the same as those in the analyses using our main sample. The observed tendency in physician decisions cannot be attributed to the end-of-shift effect.

ED Crowding.—Another potential confounding factor in estimating the decision fatigue effect is ED crowding. Physicians continuously monitor the ED queue status through a computer terminal. Studies have found that overcrowding in the ED influences not only discharge decisions, but also test ordering and patients' length of stay (Gorski et al., 2017; Freeman et al., 2017; Chiu et al., 2018). To address this concern, we conduct robustness analyses that control for ED crowding.

We have two measures for the degree of crowding in the ED. The first is the volume of patients waiting to be seen in the ED at the time of the index patient's consultation starting. The second is the real-time system load in the ED, including those waiting to be seen and those being treated.

Panels C and D in Table 4 present estimated coefficients after controlling for the congestion in the waiting area and the overall system load, respectively. Regardless of our different measurements, the point estimates on decision fatigue are essentially the same as those in our baseline estimation, suggesting that the variation in ED patient volume is not the main source of the observed physician behavioral differences.

Physicians' Multitasking.—ED physicians attend to multiple patients at the same time due to the increased demand for emergency services. However, an emerging literature in experimental psychology and cognitive neuroscience has demonstrated that multitasking impairs workers' decision making and decreases productivity (Rubinstein et al., 2001; Hallowell, 2005). In healthcare operations, a recent work of KC (2014) identifies that excessive multitasking in ED physicians adversely impacts productivity and quality of care.

Given this concern, we further account for physicians' multitasking as a robustness check. Here, we define multitasking as the number of patients concurrently managed by the physician during the index patient's consultation. Panel E of Table 4 shows that the results remain largely unchanged after controlling for the level of physician multitasking.

Patient Diagnostics.—The analyses above only control for ex ante patient characteristics of demographics and emergency severity level. As a robustness check, we control for the ex post clinical characteristics in our regressions. Clinical diagnoses, of course, are also partly determined by patient care and physician diagnostic performance. We are cautious, therefore, in interpreting these estimates. We divide patient primary diagnostic information into 20 categories based on the ICD-9-CM, and control for 19 diagnosis fixed effects in our analysis.[4] Panel F of Table 4 shows that results are essentially unchanged regardless of whether we control for clinical diagnoses.

Restrictions on Physician Shift Length.—We further examine the robustness of our results with respect to the restrictions on physician shift length. We define shift duration to be between 8 and 12 hours, between 8 and 10 hours, and of about 8 hours. Results for the subsamples, shown in Panels G to I of Table 4, are statistically significant and negative. Point estimates are similar in magnitude compared with those in Table 2.

[4] Diseases are classified into 20 categories: infectious and parasitic diseases, neoplasms, endocrine-nutritional-metabolic diseases, diseases of the blood, mental disorders, diseases of the nervous system, diseases of the 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, diseases of the skin and subcutaneous tissue, diseases of the musculoskeletal system, congenital anomalies, disorders originating in the perinatal period, signs-symptoms, injury-poisoning, external causes of injury, and supplementary classification.

E. Heterogeneous Analyses

Nonlinear Decision-Fatigue Effects.—While decision fatigue has been shown to affect physician behavior, its effects might not be a constant. To assess the extent of this heterogeneity, we re-estimate equation (1) by replacing the linear measure of decision fatigue with indicator variables for patient positions in the shift. The estimation equation is as follows:

$$(3) \qquad Y_{ijt} = \sum_m \alpha_m D_{ijm} + X_i\beta + T_t\gamma + \nu_j + \epsilon_{ijt}.$$

In this specification, we categorize patient visits into eight groups. The first group includes visits whose consultation starts when the physician has previously treated 0 to 2 patients during the shift. The second group includes visits treated after 3 to 5 patients, the third 6 to 8, and so on through 18 to 20, with a final group of more than 20. In Equation (3), D_{ijm} is a dummy variable indicating the number of patients seen by physician j before patient $i's$ consultation falls into group m ($m \in \{1, 2, 3, 4, 5, 6, 7, 8\}$). The eighth group, for visits whose physicians have previously treated more than 20 cases during the shift, is the reference category.

Table 5 shows estimation results from Equation (3), and Figure 2 plots the estimated coefficients. OLS estimates for α_m remain significant and positive for most models. As shown in Figure 2, the estimates decrease in magnitude as the grouping number increases. In particular, the reduction is sharper and more pronounced in the earlier part of a shift than in the later part. This result also rules out the possibility that physicians speed up only as they approach the end of a shift.

Severity of Previous Cases.—Previous models use the total number of patients treated as a measure of the physician's decision fatigue. However, the composition of previously treated patients might be very different: Treating ten patients with complex or severe conditions would be quite different from treating ten patients with mild illnesses. More severe cases need more physician effort in terms of concentration and mental inputs, and thus lead to a higher degree of decision fatigue.

To examine the heterogeneous effects of previous cases—i.e., the number and level of severity—we conduct an analysis that controls for both the total number of patients treated and the number of severe cases. Panel A of Table 6 reports the estimation results. All models estimate negative coefficients for these two variables. Fixing the total number of patients treated, an increase in the number of severe cases increases the effects of decision fatigue. This result is consistent with our prediction: Severe cases lead to more decision fatigue for physicians than non-severe cases.

Patient-Physician Race and Gender Concordance.—Concordance by race and gender in patient-physician relationships is associated with greater patient satisfaction, more productive communication and better exchange of health information (Cooper-Patrick et al., 1999; Street and Haidet, 2011). We then examine whether the effects of decision fatigue vary by race and gender concordance status of patients.

Panel B of Table 6 shows decision fatigue displays a smaller effect on patients who share a same race or gender as the given physician. In particular, Column (1) of Panel B suggests that the reduction in the probability of inpatient admission caused by increased decision fatigue is significantly smaller when the index patient has the same gender as the physician. Column (2) shows that the reduction in the number of task orders is alleviated by both race and gender concordance between patients and physicians.

These observations could be possibly driven by two channels. First, physicians have a better knowledge of their race (or gender) concordant patients. Second, shared identities facilitate more productive patient-physician communication. Both channels suggest that physicians could gather diagnostic information at a lower cost of mental capacity. As a result, decision fatigue displays a relatively smaller effect on physicians' race (or gender) concordant patients.

Severe vs. Non-Severe Current Cases.—To investigate whether decision fatigue affects physicians' decisions differently when they treat patients of different severity levels, we estimate separate models for severe and non-severe patient visits. As suggested in Panels C and D of Table 6, OLS estimates remain statistically significant and negative for these two subsamples. For both groups, physicians reduce the number of inpatient admissions, issue fewer task orders, and shorten patient length of stay as they treat more patients during a shift. However, the effect size is larger for non-severe cases than severe cases in terms of percentage changes. One explanation might be that physicians are more alert and focused when treating severe cases, which would reduce the effects of mental fatigue.

IV. Impact of Physician Decision Fatigue on Patient Outcomes

In this section, we examine whether decision fatigue worsens patient health outcomes. We focus on two patient outcomes: Whether a patient revisited the ED within two weeks, and whether the patient died in the ED after arrival.

For the purpose of illustration, we stratify patient visits into two groups based on their sequences in the shift. On average, a physician has treated 10.4 patients before the index patient's consultation in the shift. The first group comprises the first 11 patients to visit the physician and the second group the remaining patients in a given shift. Panel A in Figure 3 depicts 14-day ED revisit rates for these two groups. Since ED death occurs only in severe cases, Panel B compares the mortality rates among severe cases for these two groups. Figure 3 shows that patients who arrive later in the shift have higher probabilities of ED revisits and death. Specifically, the revisit rate increases from 12.1% for the first group to 13.9% for the second group, and the mortality rate from 0.8% to 1%. The differences are statistically significant at the 1% level for both outcomes.

To rigorously examine the relationship between physician decision fatigue and patient outcomes, we conduct regression analyses that control for patient demographics, case severity, physician fixed effects, and a series of time fixed effects. The regression specification is the same as Equation (1). For patient revisits, outcome variable Y_{ijt} is a dummy variable that equals one if patient i revisited the ED within two weeks and zero otherwise. For mortality, the outcome variable is a dummy variable that equals one if the patient died in the ED after arrival and zero otherwise.

Table 7 shows that physician decision fatigue has statistically significant and positive effects on both 14-day revisits to the ED and mortality in the ED. As shown in Column (1), every ten patients the physician has previously treated during a shift increase the index patient's revisit probability by 0.46 percentage point, which is 3.6% more than a sample mean likelihood of 0.129. Column (2) suggests that every ten patients previously treated are associated with a 0.03 percentage point (or 12.5% of the sample mortality rate) increase in the probability of death in the ED.

Our results imply that physician decision fatigue may increase patient health risk, leading to higher likelihood of ED revisits and even mortality. Revisits are expensive and often more costly than initial visits (Duseja et al., 2015). Based on our data, we conservatively estimate a cost of $200 per revisit. Then if all patients had been treated after the physician had previously seen ten more patients in the shift, decision fatigue would have cost the ED approximately of $130,000 extra per year.

V. Medical Experience and Decision Fatigue

Results in the sections above indicate that decision fatigue affects physician decisions and has a negative effect on patient health. Based on the observation that increased experience reduces cognitive workload (Patten et al., 2006), here we examine whether and how medical experience reduces the effects of decision fatigue on physicians' decision-making.

We first estimate the decision-fatigue effect for each physician and subsequently regress this estimate on the physician's characteristics. The equation that characterizes the relationship between decision-fatigue effects and medical experience is:

$$(4) \qquad \alpha_{Yj} = f(exp_j, X_j) + \epsilon_{Yj},$$

where the dependent variable α_{Yj} is the IV estimate of individual-specific decision-fatigue effect on decision Y for physician j. The variable of interest, exp_j, measures years of medical experience for physician j. X_j represents other physician characteristics, including gender, one dummy variable that indicates an initial degree from a local medical school, and another dummy variable that indicates continuing medical education. ϵ_{Yj} is the error term. The function $f(exp_j, X_j)$ represents the relationship between decision-fatigue effects and medical experience conditional on covariates X_j. Below, we provide three specifications for $f(exp_j, X_j)$ to conduct the estimation.

We start with a quadratic model. In this specification, medical experience enters Equation (4) in a quadratic form, such that:

$$(5) \qquad \alpha_{Yj} = \beta_0 + \beta_1\, exp_j + \beta_2\, exp_j^2 + X_j\gamma + \epsilon_{Yj}.$$

Table 8 presents OLS estimation results. As shown in the first two rows, the estimate of β_1 is statistically significant and positive, and the estimate of β_2 is also statistically significant but negative in all columns. This result suggests an inverted U-shaped relationship between decision-fatigue effects and medical experience. More specifically, decision fatigue exerts less influence when physicians' experience is moderate rather than high or low.

We further use a cubic spline regression to plot the relationship between decision-fatigue effects and medical experience. The splines provide great flexibility for fitting data. The dashed gray line in Figure 4 depicts the fitted values, $\hat{\alpha}$, from the smoothing splines. As predicted, the fitted relationship exhibits a nonlinear pattern. The graph has a positive slope from low to moderate values for experience. However, the curve becomes flatter, or even turns negative, when years of experience continue to increase.

Finally, we perform a two-lines test to examine the inverted U-shaped relationship (Simonsohn, 2018). This newly proposed test involves estimating two regression lines, one for low and one for high values of the x-axis variable. A significant U-shape exists if the coefficients of these two lines have opposite signs and are individually statistically significant.

Figure 4 plots the average slopes for low and high values of experience using the two-lines test. The first lines have statistically significant and positive slopes in all panels. In contrast, the second lines have flat or negative slopes, suggesting that increasing medical experience exhibits little or even negative effect for physicians with moderate to high values for experience. Moreover, we find a significantly inverted U-shape in Panel C, and no significant U-shapes in Panels A and B. In summary, this indicates that medical experience mitigates decision-fatigue effects for young physicians.

Taken together, our results suggest that professional experience mitigates decision-fatigue effects for young physicians, while this effect fades as the physician ages.

VI. Conclusion

Using administrative data of 240,000 ED visits, we find that decision fatigue erodes the quality of treatment provided by the physician and impairs patient outcomes. Increased decision fatigue of physicians leads to lower inpatient admission rates, fewer task orders, and shorter patient length of stay; subsequently, both patient revisit rates and mortality rates increase. Our results also show that a break in the shift and physician's medical experience could alleviate the consequences of decision fatigue. Researchers have initiated efforts to design "choice architecture" or "nudges" to improve the quality of medical decision-making (Avorn, 2018). In this regard, to reduce decision fatigue, hospitals could introduce more breaks or shorter shifts, assign serious cases to less fatigued physicians, and delegate certain decisions to support systems and subordinates (Berner and Graber, 2008).

Our study has broad implications for public debates on regulations for healthcare professionals. Though strict occupational licensing may ensure the quality of physicians, it also contributes to staffing shortages in the healthcare industry that cause excessive workloads for physicians (Darzi and Evans, 2016). Setting a high benchmark for qualifying physicians may have the unintended consequence of increasing the likelihood of physician shortages, greater fatigue, and degraded quality of treatment. We suggest that there is a quid pro quo incurred by demanding higher qualifications and fewer healthcare professionals that policy makers must consider in light of our findings on the salient features of physician fatigue that could result in degraded medical decision-making (Friedman, 2009).

REFERENCES

Apker, Julie, Larry A Mallak, and Scott C Gibson. 2007. "Communicating in the "Gray Zone": Perceptions about Emergency Physician–Hospitalist Handoffs and Patient Safety." *Academic Emergency Medicine*, 14(10): 884–894.

Augenblick, Ned, and Scott Nicholson. 2015. "Ballot Position, Choice Fatigue, and Voter Behaviour." *The Review of Economic Studies*, 83(2): 460–480.

Avorn, Jerry. 2018. "The Psychology of Clinical Decision Making—Implications for Medication Use." *The New England Journal of Medicine*, 378(8): 689–691.

Baumeister, Roy F, Ellen Bratslavsky, Mark Muraven, and Dianne M Tice. 1998. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology*, 74(5): 1252.

Berner, Eta S., and Mark L Graber. 2008. "Overconfidence as a Cause of Diagnostic Error in Medicine." *The American Journal of Medicine*, 121(5): S2–S23.

Blendon, Robert J, Catherine M DesRoches, Mollyann Brodie, John M Benson, Allison B Rosen, Eric Schneider, Drew E Altman, Kinga Zapert, Melissa J Herrmann, and Annie E Steffenson. 2002. "Views of Practicing Physicians and the Public on Medical Errors." *The New England Journal of Medicine*, 347(24): 1933–1940.

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.

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

Chan, David C. 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." *Handbook of Health Economics*, 2 397–432.

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.

Chiu, I-Min, Yan-Ren Lin, Yuan-Jhen Syue, Chia-Te Kung, Kuan-Han Wu, and Chao-Jui Li. 2018. "The Influence of Crowding on Clinical Practice in The Emergency Department." *American Journal of Emergency Medicine*, 36(1): 56–60.

Clemens, Jeffrey, and Joshua D Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review*, 104(4): 1320–49.

Cohen, I Glenn, Holly Fernandez Lynch, and Christopher T Robertson. 2016. *Nudging Health: Health Law and Behavioral Economics*. Johns Hopkins University Press.

Cooper-Patrick, Lisa, Joseph J Gallo, Junius J Gonzales, Hong Thi Vu, Neil R Powe, Christine Nelson, and Daniel E Ford. 1999. "Race, Gender, and Partnership in the Patient-Physician Relationship." *JAMA*, 282(6): 583–589.

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

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

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

Darzi, Ara, and Tim Evans. 2016. "The Global Shortage of Health Workers—An Opportunity to Transform Care." *Lancet*, 388(10060): 2576–2577.

Duseja, Reena, Naomi S Bardach, Grace A Lin, Jinoos Yazdany, Mitzi L Dean, Theodore H Clay, W John Boscardin, and R Adams Dudley. 2015. "Revisit Rates and Associated Costs After an Emergency Department Encounter: A Multistate Analysis." *Annals of Internal Medicine*, 162(11): 750–756.

Dzau, Victor J, Darrell G Kirch, and Thomas J Nasca. 2018. "To Care Is Human—Collectively Confronting the Clinician-Burnout Crisis." *The New England Journal of Medicine*, 378(4): 312–314.

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

Freeman, Michael, Susan Robinson, and Stefan Scholtes. 2017. "Gatekeeping under Congestion: An Empirical Study of Referral Errors in the Emergency Department." https://ssrn.com/abstract=3036999.

Friedman, Milton. 2009. *Capitalism and Freedom*. University of Chicago Press.

Gee, Alastair. 2011. "Concerns Raised over New US Resident Physician Work Hours." *Lancet*, 378(9793): 758.

Gorski, Jillian K, Robert J Batt, Erkin Otles, Manish N Shah, Azita G Hamedani, and Brian W Patterson. 2017. "The Impact of Emergency Department Census on the Decision to Admit." *Academic Emergency Medicine*, 24(1): 13–21.

Hallowell, EM. 2005. "Overloaded Circuits: Why Smart People Underperform." *Harvard Business Review*, 11(3): 289–299.

Hirshleifer, David, Yaron Levi, Ben Lourie, and Siew Hong Teoh. Forthcoming 2018. "Decision Fatigue and Heuristic Analyst Forecasts." *Journal of Financial Economics*.

Johnson, Eric J. 2008. "Man, My Brain is Tired: Linking Depletion and Cognitive Effort in Choice." *Journal of Consumer Psychology*, 18(1): 14–16.

Kahneman, Daniel, and Patrick Egan. 2011. *Thinking, Fast and Slow,* Vol. 1. New York: Farrar, Straus and Giroux.

KC, Diwas Singh. 2014. "Does Multitasking Improve Performance? Evidence from the Emergency Department." *Manufacturing & Service Operations Management*, 16(2): 168–183.

Levav, Jonathan, Mark Heitmann, Andreas Herrmann, and Sheena S Iyengar. 2010. "Order in Product Customization Decisions: Evidence from Field Experiments." *Journal of Political Economy*, 118(2): 274–299.

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.

Linder, Jeffrey A, Jason N Doctor, Mark W Friedberg, Harry Reyes Nieva, Caroline Birks, Daniella Meeker, and Craig R Fox. 2014. "Time of Day and the Decision to Prescribe Antibiotics." *JAMA Internal Medicine,* 174(12): 2029–2031.

Lu, Susan F, and Huaxia Rui. 2017. "Can We Trust Online Physician Ratings? Evidence from Cardiac Surgeons in Florida." *Management Science*, 64(6): 2557–2573.

Makary, Martin A, and Michael Daniel. 2016. "Medical Error—the Third Leading Cause of Death in the US." *BMJ*, 353, p. i2139.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. "Poverty Impedes Cognitive Function." *Science*, 341(6149): 976–980.

Michtalik, Henry J, Hsin-Chieh Yeh, Peter J Pronovost, and Daniel J Brotman. 2013. "Impact of Attending Physician Workload on Patient Care: A Survey of Hospitalists." *JAMA Internal Medicine*, 173(5): 375–377.

Morrow, Gill, Bryan Burford, Madeline Carter, and Jan Illing. 2014. "Have Restricted Working Hours Reduced Junior Doctors' Experience of Fatigue? A Focus Group and Telephone Interview Study." *BMJ*, 4(3): e004222.

Mullainathan, Sendhil and Eldar Shafir. 2013. *Scarcity: Why Having Too Little Means So Much*, Macmillan.

Nasca, Thomas J, Susan H Day, and E Stephen Amis Jr. 2010. "The New Recommendations on Duty Hours from The ACGME Task Force." *The New England Journal of Medicine*, 363, p. e3.

Patten, Christopher JD, Albert Kircher, Joakim Östlund, Lena Nilsson, and Ola Svenson. 2006. "Driver Experience and Cognitive Workload in Different Traffic Environments." *Accident Analysis & Prevention*, 38(5): 887–894.

Philibert, Ingrid, Paul Friedmann, William T Williams et al. 2002. "New Requirements for Resident Duty Hours." *JAMA*, 288(9): 1112–1114.

Physicians Foundation. 2016. A Survey of America's Physicians: Practice Patterns and Perspectives. *The Physicians Foundation*.

Rubinstein, Joshua S, David E Meyer, and Jeffrey E Evans. 2001. "Executive Control of Cognitive Processes in Task Switching." *Journal of Experimental Psychology: Human Perception and Performance*, 27(4): 763-797.

Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir. 2012. "Some Consequences of Having Too Little." *Science*, 338(6107): 682–685.

Sievertsen, Hans Henrik, Francesca Gino, and Marco Piovesan. 2016. "Cognitive Fatigue Influences Students' Performance on Standardized Tests." *Proceedings of the National Academy of Sciences*, 113(10): 2621–2624.

Simonsohn, U. 2018. "Two-Lines: A Valid Alternative to The Invalid Testing of U-Shaped Relationships with Quadratic Regressions." https://ssrn.com/abstract=3021690.

Skinner, Jonathan. 2011. "Causes and Consequences of Regional Variations in Health Care." *Handbook of Health Economics*, Vol 2, 45–93.

Slovic, Paul, Melissa L Finucane, Ellen Peters, and Donald G MacGregor. 2007. "The Affect Heuristic." *European Journal of Operational Research*, 177(3): 1333–1352.

Street, RJ, and P Haidet. 2011. "How Well Do Doctors Know Their Patients? Factors Affecting Physician Understanding of Patients' Health Beliefs." *Journal of General Internal Medicine*, 26(1): 21–27.

Tyler, James M, and Kathleen C Burns. 2008. "After Depletion: The Replenishment of the Self's Regulatory Resources." *Self Identity*, 7(3): 305–321.

Vohs, KD, RF Baumeister, BJ Schmeichel, JM Twenge, NM Nelson, and DM Tice. 2008. "Making Choices Impairs Subsequent Self-Control: A Limited-Resource Account of Decision Making, Self-Regulation, and Active Initiative." *Journal of Personality and Social Psychology*, 94(5): 883–898.

Wakker, Peter P. 2008. "Lessons Learned By (From?) An Economist Working in Medical Decision Making." *Medical Decision Making*, 28(5): 690–698.

West, Colin P, Liselotte N Dyrbye, Patricia J Erwin, and Tait D Shanafelt. 2016. "Interventions to Prevent and Reduce Physician Burnout: A Systematic Review and Meta-Analysis." *Lancet*, 388(10057): 2272–2281.

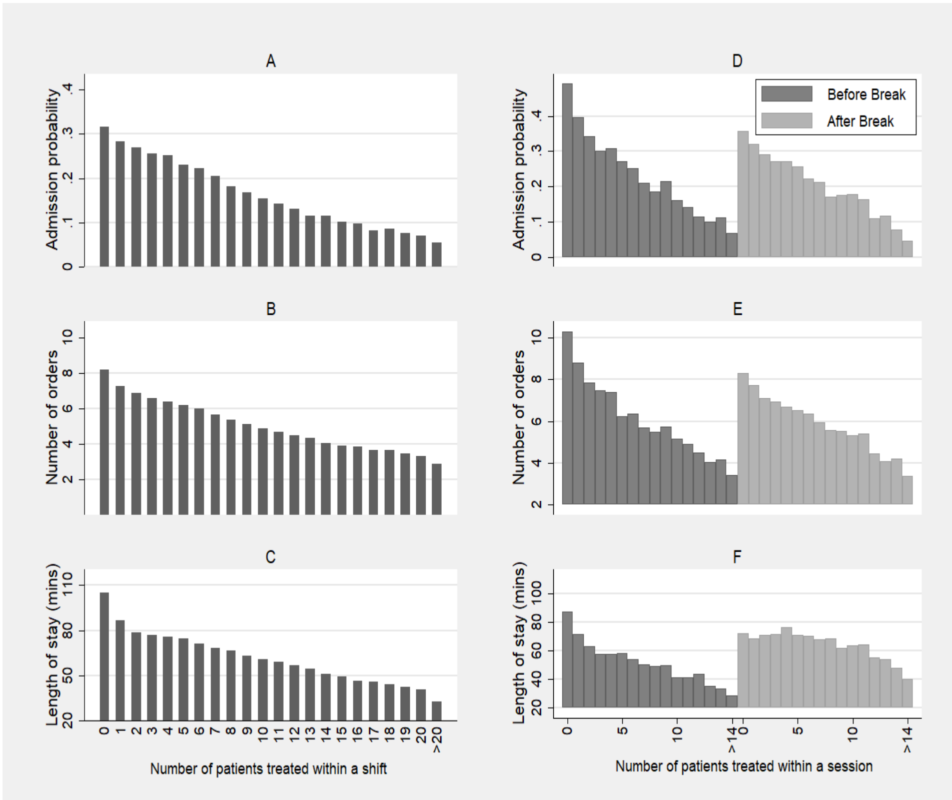

FIGURE 1. PHYSICIAN DECISIONS BY NUMBER OF PATIENTS TREATED WITHIN A SHIFT OR A SESSION

Notes: In these panels, the y-axis plots the mean values of three measures of physician decisions for the index patient: proportion of hospital admissions, number of task orders, and patient length of stay (in minutes). In Panels A-C, the x-axis represents the number of patients that physicians have seen previously during the same shift. In Panels D-E, the x-axis shows the number of patients that physicians have seen previously during each of the two sessions divided by a break. Here a break is defined as a period of at least one hour, during which the physician on shift is not in charge of any patient. Panels D-E only include shifts with one break.

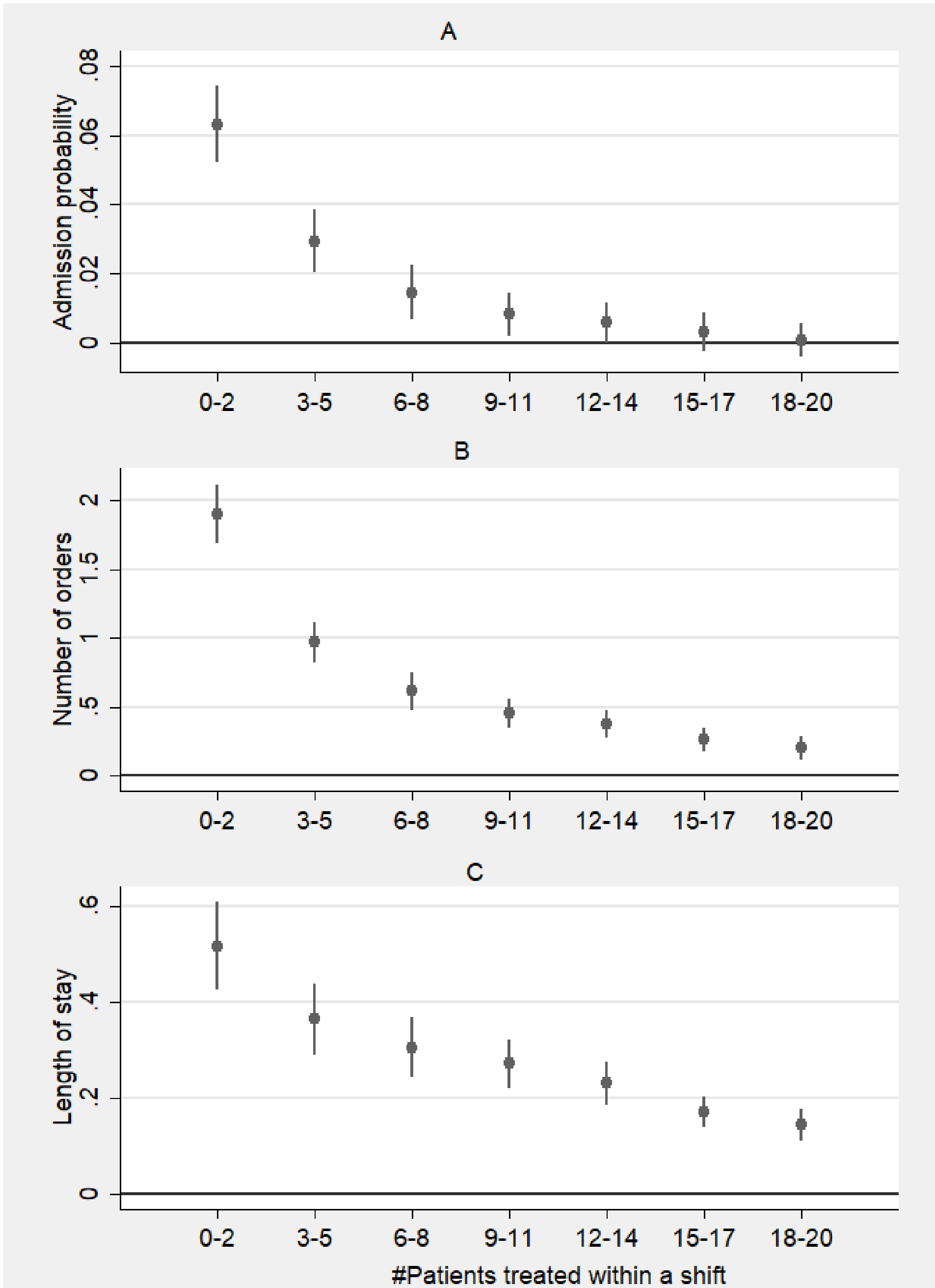

FIGURE 2. HETEROGENEOUS ANALYSIS—NONLINEAR DECISION-FATIGUE EFFECTS

Notes: This figure plots coefficients from Equation (5) estimated separately for hospital admission (Panel A), number of task orders (Panel B), and (log) length of stay (Panel C). The reference category includes patient visits whose consultation starts when the physician has treated more than 20 patients during the shift. Dots represent point estimates from regression models, and solid bars represent the 95% confidence interval for each estimate. Results are also presented in Table 5.

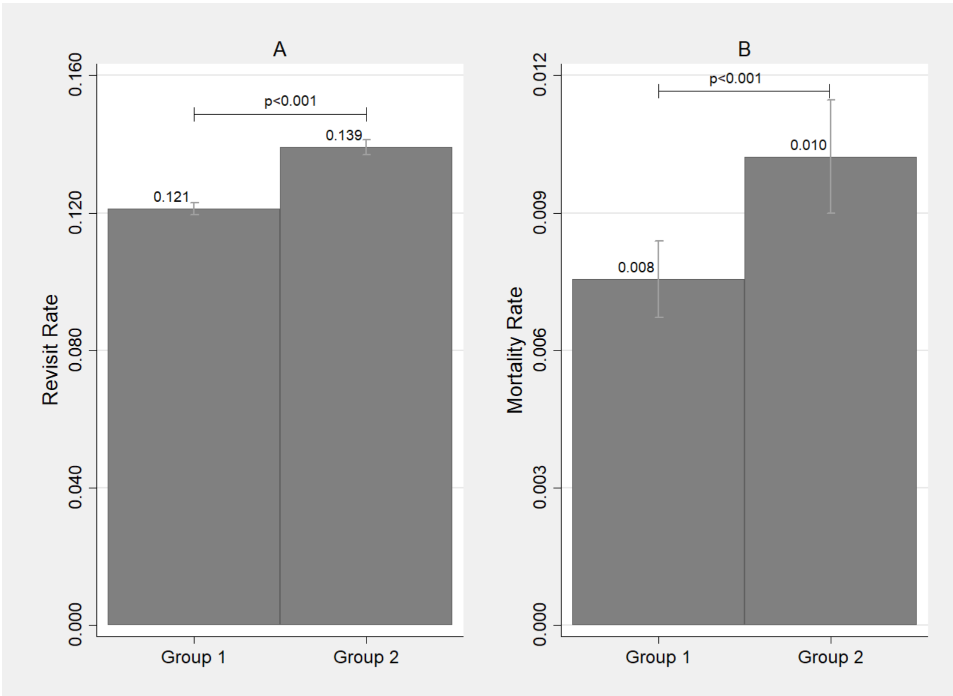

FIGURE 3. PATIENT TREATMENT OUTCOMES BY ORDINAL POSITION

Notes: Patients are stratified into two groups by their ordinal position in the shift. Group 1 comprises patients who were among the first 11 to visit the physician, and Group 2 the remaining patients. We compare the two groups in terms of the rates of revisits within two weeks in Panel A, and the rates of mortality among severe cases in Panel B. Error bars represent 95% CI. The p values reported above the top horizontal bars are from Chi-squared tests of differences in means between Groups 1 and 2.

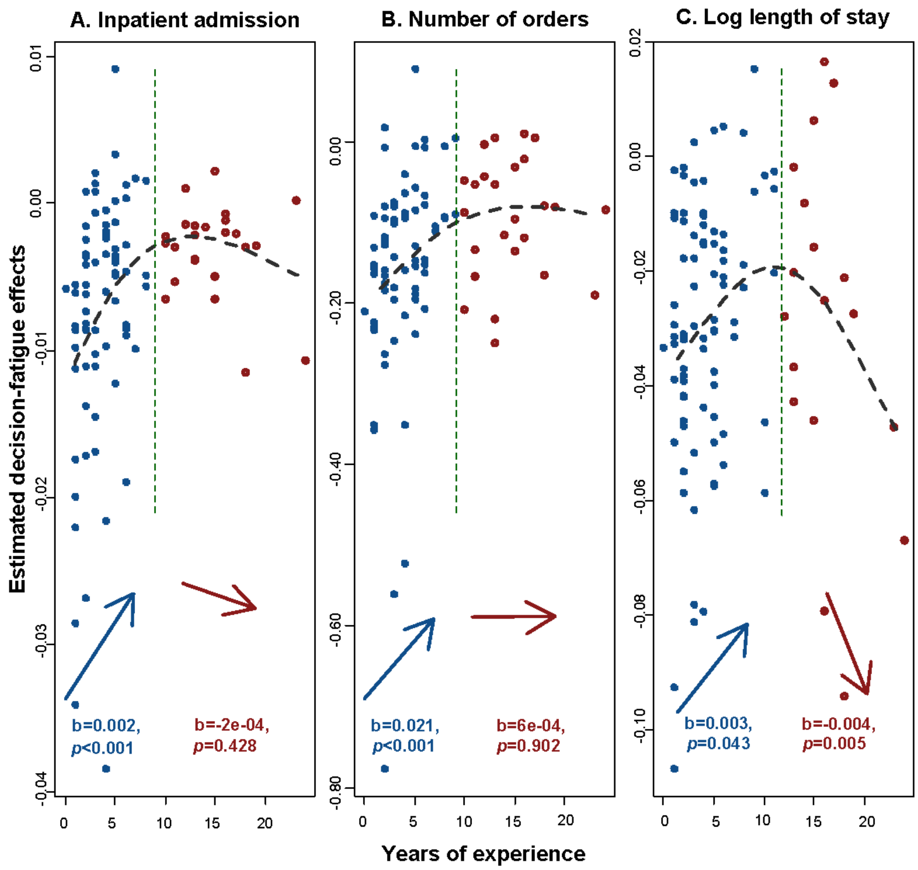

FIGURE 4. EFFECTS OF DECISION FATIGUE BY PHYSICIAN'S EXPERIENCE

Notes: This figure shows the association between years of experience and decision-fatigue effects on inpatient admission (Panel A), number of orders (Panel B), and log length of stay (Panel C). Each circle represents a physician in the sample. The dashed gray line depicts fitted relationships between decision-fatigue effect and medical experience, using cubic splines. Arrows plot the average slopes for low and high values of experience using a newly proposed test for nonlinearity (Simonsohn, 2018). All models control for the covariates of physician's gender, country of graduation, and acquisition of continuing medical education.

TABLE 1—SUMMARY STATISTICS

Variable	Observations	Mean	SD
Panel A: Physician decision fatigue			
Number of patients treated [a]	242,761	10.350	8.334
Panel B: Physician decisions			
Inpatient admission	242,761	0.176	0.381
Total number of orders	204,510	5.317	5.021
Patient length of stay, minutes	242,753	63.356	73.598
Panel C: Patient outcomes			
Return visits within 14 days	242,761	0.129	0.335
Death in the ED	242,761	0.002	0.048
Panel D: Patient characteristics			
Male	242,761	0.648	0.478
Age	242,761	39.184	20.433
Patient severity level			
1	242,761	0.038	0.192
2	242,761	0.237	0.425
3	242,761	0.725	0.447
Panel E: Physician characteristics			
Years of experience	101	6.406	5.398
Male	101	0.772	0.421
Local graduates	101	0.515	0.502
Continuing medical education	101	0.248	0.434
Panel F: IV			
Number of ambulance arrivals [b]	242,761	9.754	7.759

[a] Number of patients treated by the physician from the shift's start to the index patient's consultation.

[b] Total number of ambulance arrivals at the ED during the physician's shift up to the index patient's consultation.

TABLE 2— EFFECTS OF DECISION FATIGUE ON PHYSICIAN BEHAVIOR

	(1) Inpatient admission	(2) Inpatient admission	(3) Number of orders	(4) Log length of stay
Panel A	**OLS**	**Probit**	**OLS**	**OLS**
# Patients treated	-0.0021***	-0.0029***	-0.0655***	-0.0191***
	(0.0002)	(0.0004)	(0.0054)	(0.0018)
R-squared	0.378	0.376	0.515	0.329
Percent effect: #Patients treated +10	11.9	16.5	12.3	19.1
Panel B: First stage results				
#Ambulance arrivals	0.5573***	0.5573***	0.5415***	0.5573***
	(0.0230)	(0.0230)	(0.0222)	(0.0230)
R-squared	0.641	0.641	0.636	0.641
Panel C: IV regressions				
# Patients treated	-0.0034***	-0.0037***	-0.1014***	-0.0230***
	(0.0004)	(0.0005)	(0.0082)	(0.0022)
R-squared	0.377	-	0.513	0.329
Percent effect: #Patients treated +10	19.3	20.8	19.1	23.0
P value of Hausman test	<0.001	<0.001	<0.001	<0.001
Patient characteristics	YES	YES	YES	YES
Physician FE	YES	YES	YES	YES
Time FE	YES	YES	YES	YES
Observations	242761	242761	204510	242752
Sample mean outcome	0.176	0.176	5.317	3.387

Notes: Physician decision fatigue is measured by the number of patients treated before an index patient's consultation. Panel A reports OLS/probit estimates. Panel B reports first-stage estimates over the sample that is used in the IV regressions in Panel C. Panel C reports IV estimates, in which the instrumental variable is the number of hospital ambulance arrivals during the physician's shift up to the index patient's consultation. Dependent variables are a dummy variable that equals one if the patient is admitted and zero otherwise (Column (1) and (2)), total number of task orders (Column (3)), and patient length of stay in logarithmic form (Column (4)). All regressions control for patient demographic characteristics (gender, age, age squared, and race); case severity; time fixed effects (hour of day, day of week, and month-year interactions); and physician fixed effects. Standard errors in parentheses are clustered at the physician level.

[a] Panels A and C in Column (2) report average marginal effects from the probit model.

[b] Percentage changes in the dependent variable relative to the mean, when the number of previously treated patients increases by 10 units. For example, every ten patients the physician had previously treated during a shift reduced the index patient's inpatient admission probability by 2.1 percentage points, which translates an 11.9% (2.1/17.6) reduction.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 3—BREAK WITHIN THE SHIFT

	(1)	(2)	(3)
	Inpatient admission	Number of orders	Log length of stay
Group 1 (α1)	0.0118	0.2887*	0.1582***
	(0.0134)	(0.1538)	(0.0352)
Group 3 (α2)	0.0260*	1.1302***	0.2235***
	(0.0139)	(0.1958)	(0.0461)
α1+α2	0.0378	1.4189	0.3817
P value of Wald test	0.0937	0.0000	0.0000
Patient characteristics	YES	YES	YES
Physician FE	YES	YES	YES
Time FE	YES	YES	YES
Observations	7, 285	6,458	7, 285
R-squared	0.387	0.504	0.426
Sample mean outcome	0.279	7.026	3.367

Notes: This table reports coefficient estimates from OLS regressions using Equation (2). Based on administrative data, we extract shifts with a break during which the physician is not in charge of any patient for one hour or more. We focus on three groups of patient visits in the extracted shifts. Group 1 refers to the last fourth to sixth cases before the break, Group 2 the last three cases before the break, and Group 3 the first three cases after the break. Group 2 is used as the omitted group in the regression, Outcome variables and controls are the same as in Table 2. Standard errors in parentheses are clustered at the physician level.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 4—ROBUSTNESS ANALYSES

	(1) Inpatient admission	(2) Number of orders	(3) Log length of stay
Panel A: control for cumulative time elapsed in the shift			
# Patients treated	-0.0010***	-0.0342***	-0.0127***
	(0.0002)	(0.0049)	(0.0021)
Panel B: exclude the last three visits in each shift			
# Patients treated	-0.0023***	-0.0607***	-0.0150***
	(0.0003)	(0.0059)	(0.0019)
Panel C: control for ED patients-in-waiting			
# Patients treated	-0.0021***	-0.0647***	-0.0192***
	(0.0002)	(0.0053)	(0.0018)
Panel D: control for ED system load			
# Patients treated	-0.0021***	-0.0643***	-0.0195***
	(0.0002)	(0.0053)	(0.0019)
Panel E: control for physician multitasking			
# Patients treated	-0.0021***	-0.0643***	-0.0182***
	(0.0002)	(0.0059)	(0.0018)
Panel F: control for patient diagnostics			
# Patients treated	-0.0020***	-0.0606***	-0.0180***
	(0.0002)	(0.0051)	(0.0018)
Panel G: restrictions on physician shift length (8-12 hours)			
# Patients treated	-0.0020***	-0.0632***	-0.0206***
	(0.0002)	(0.0054)	(0.0016)
Panel H: restrictions on physician shift length (8-10 hours)			
# Patients treated	-0.0016***	-0.0515***	-0.0199***
	(0.0002)	(0.0048)	(0.0016)
Panel I: restrictions on physician shift length (8 hours)			
# Patients treated	-0.0013***	-0.0342***	-0.0179***
	(0.0002)	(0.0041)	(0.0023)

Notes: This table reports coefficients from OLS regressions using Equation (1). Panel A adds the number of hours elapsed in the given shift as a control variable. Panel B excludes the last three visits in each shift. Panel C controls for the number of patients waiting to be seen in the ED. Panel D controls for the total number of patients in the ED. Panel E controls for the physician's degree of multitasking, measured by the number of overlapping cases. Panel F controls for patient diagnostic indicators. Panels G to I restrict the shift length to be between 8 and 12 hours, between 8 and 10 hours, and of 8 hours, respectively. Outcome variables and other controls are the same as in Table 2. Standard errors in parentheses are clustered at the physician level.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 5—NONLINEAR DECISION-FATIGUE EFFECTS

	(1) Inpatient admission	(2) Number of orders	(3) Log length of stay
# Patients treated:			
0-2	0.0633***	1.9045***	0.5178***
	(0.0056)	(0.1088)	(0.0459)
3-5	0.0294***	0.9679***	0.3639***
	(0.0046)	(0.0754)	(0.0372)
6-8	0.0145***	0.6127***	0.3057***
	(0.0040)	(0.0683)	(0.0311)
9-11	0.0082**	0.4501***	0.2720***
	(0.0031)	(0.0558)	(0.0258)
12-14	0.0058*	0.3751***	0.2307***
	(0.0030)	(0.0518)	(0.0224)
15-17	0.0030	0.2575***	0.1709***
	(0.0028)	(0.0419)	(0.0167)
18-20	0.0007	0.1976***	0.1439***
	(0.0025)	(0.0438)	(0.0168)
Patient characteristics	YES	YES	YES
Physician FE	YES	YES	YES
Time FE	YES	YES	YES
Observations	242,761	204,510	242,752
R-squared	0.379	0.519	0.329
Sample mean outcome	0.176	5.317	3.387

Notes: This table reports coefficients from OLS regressions using Equation (3). Results are graphically shown in Figure 2. Outcome variables and controls are the same as in Table 2. Standard errors in parentheses are clustered at the physician level.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 6—HETEROGENEOUS ANALYSES

	(1) Inpatient admission	(2) Number of orders	(3) Log length of stay
Panel A: number of severe cases previously treated			
#Patients treated	-0.0019***	-0.0581***	-0.0187***
	(0.0002)	(0.0056)	(0.0016)
#Severe cases treated	-0.0036***	-0.1082***	-0.0077
	(0.0009)	(0.0204)	(0.0079)
Panel B: patient-physician race and gender concordance			
#Patients treated	-0.0023***	-0.0746***	-0.0183***
	(0.0003)	(0.0064)	(0.0029)
#Patients treated*Race-Concordance	-0.0001	0.0110***	0.0004
	(0.0001)	(0.0029)	(0.0013)
#Patients treated*Gender-Concordance	0.0006**	0.0083***	-0.0020
	(0.0003)	(0.0028)	(0.0026)
Panel C: current severe visits only			
# Patients treated	-0.0029***	-0.1264***	-0.0159**
	(0.0006)	(0.0215)	(0.0062)
percent effect: #Patients treated +10	5.8	12.7	15.9
Panel D: current non-severe visits only			
# Patients treated	-0.0022***	-0.0581***	-0.0207***
	(0.0002)	(0.0050)	(0.0016)
percent effect: #Patients treated +10	41.3	18.7	20.7

Notes: This table reports coefficients from OLS regressions using Equation (1). Panel A adds the number of severe cases treated as a control variable. Panel B controls for patient-physician race and gender concordance status, and interacts the number of patients treated with race (gender) concordance, where Race-Concordance (Gender-Concordance) is an indicator whether or not the index patient has the same race (gender) as the physician. Panel C regresses on a sample of severe visits. Panel D uses a sample of non-severe visits. Outcome variables and other controls are the same as in Table 2. Standard errors in parentheses are clustered at the physician level.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 7—PHYSICIAN DECISION FATIGUE ON PATIENT OUTCOMES

	(1)	(2)
	14-day revisit to the ED	Death in the ED
# Patients treated	0.00046***	0.00003*
	(0.00013)	(0.00002)
percent effect: #Patients treated +10	3.58	12.71
Patient characteristics	YES	YES
Physician FE	YES	YES
Time FE	YES	YES
Observations	242,761	242,761
R-squared	0.020	0.059
Sample mean outcome	0.1286	0.0024

Notes: Coefficients from OLS regressions of the number of patients treated on patient outcomes are displayed. Dependent variables in Columns (1) and (2) are dummy variables for 14-day revisit to the ED and death in the ED, respectively. Controls are the same as in Table 2. Standard errors in parentheses are clustered at the physician level.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 8—CORRELATES OF DECISION-FATIGUE EFFECTS AND PHYSICIAN CHARACTERISTICS

	Estimated decision-fatigue effects on:		
	(1)	(2)	(3)
	Inpatient admission	Number of orders	Log length of stay
Experience	0.0018***	0.0219***	0.0050***
	(0.0005)	(0.0080)	(0.0017)
Experience-squared	-0.0001***	-0.0008**	-0.0002***
	(0.0000)	(0.0004)	(0.0001)
Male	-0.0034*	-0.0377	-0.0020
	(0.0018)	(0.0288)	(0.0061)
Local graduation	0.0028*	0.0423*	-0.0008
	(0.0014)	(0.0232)	(0.0049)
Continuing medical education	0.0004	-0.0089	-0.0097
	(0.0021)	(0.0337)	(0.0072)
Observations	101	101	101
R-squared	0.196	0.135	0.095
Sample mean outcome	-0.006	-0.139	-0.029

Notes: Coefficients from OLS regressions using Equation (5). Physician-specific decision-fatigue effects are obtained from IV estimations for each physician in our sample. Outcomes in Columns (1)-(3) represent estimated decision-fatigue effects on inpatient admission, number of orders, and log length of stay, respectively.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Appendix

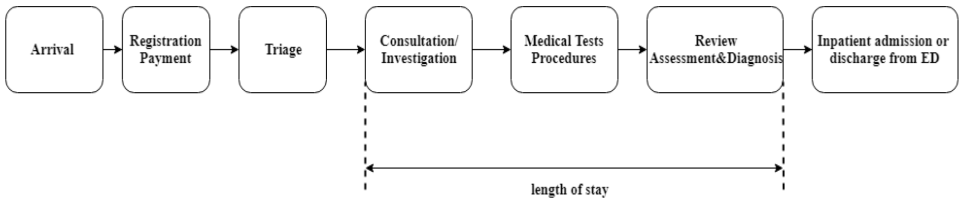

FIGURE A.1. PATIENT FLOW IN THE EMERGENCY DEPARTMENT

Notes: This figure depicts the general patient flow in the ED, starting with patient arrival and ending with the patient's being admitted to the hospital or discharged from the ED. Patient length of stay is measured from the start time of patient's consultation to the end of the consultation. Case end is not necessarily the same as consultation end. For example, a patient's consultation ends but the patient is still waiting to be admitted; the case does not end until the patient is admitted.

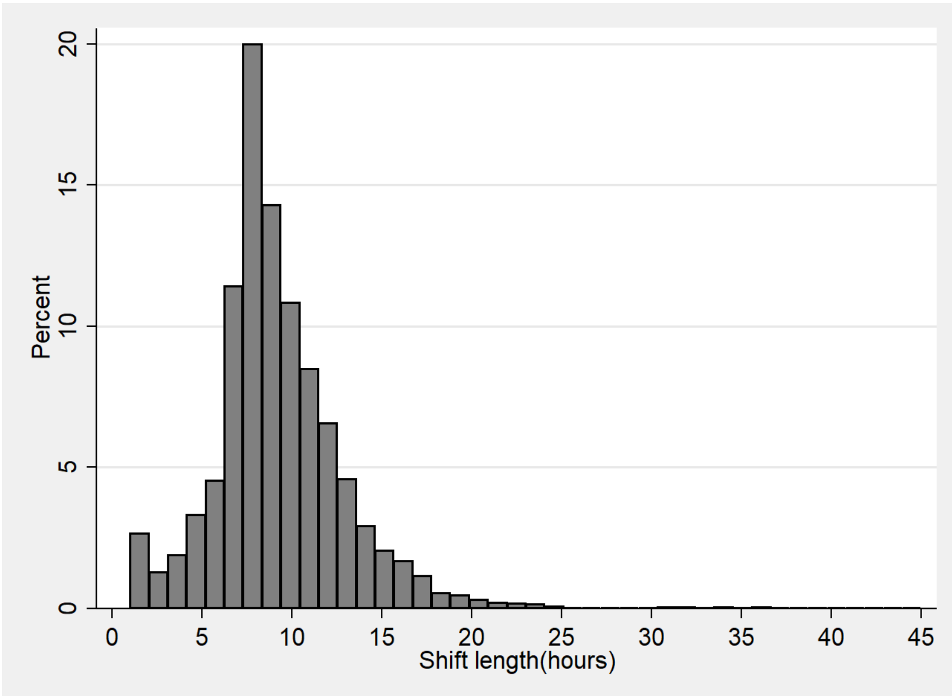

FIGURE A.2. DISTRIBUTION OF SHIFT LENGTHS

Notes: A new shift is defined to begin when a physician starts the first patient visit after six or more hours of inactivity. Shift length is measured as the number of hours elapsed from the start of the first visit's consultation to the end of the last consultation in the shift.

TABLE A.1—STATISTICS FOR AMBULANCE OPERATORS

Ambulance Operators	Frequency	Percent
SCDF Ambulance	32,129	93.41%
Private Ambulance	2,156	6.27%
Police Vehicle	39	0.11%
Others	73	0.21%
Total	34,397	100.00%

Notes: Data from patient visits sent by ambulances in the main analytic sample.