The Early Bird Gets to Live? Investigating the Relationship Between Surgery Timing and Mortality

Abstract

Fatigue impedes decision making, concentration, and overall performance. In this study, we sought to investigate whether elective surgeries performed later in the day are associated with increased patient mortality, given that doctors experience heightened fatigue at the end of the work day. Using data gathered by Sessler et.al. of 32,001 elective surgeries at the Cleveland Clinic, we fit a logistic regression model to investigate the relationship between surgery timing and patient mortality while controlling for potential confounding variables. Through our analysis, we identified statistically significant increased odds of mortality for surgeries performed later in the day when compared to surgeries performed early in the morning. These findings may motivate future hospital policy changes, such as a restructuring of surgery scheduling to earlier in the day or increased breaks, and may also inform future research on methods to mitigate physician fatigue.

I. Background and Significance

Previous literature has documented the detrimental effect of fatigue on performance—to put it simply, people make more mistakes when they're tired (1). Consequently, many worry about the impact of fatigue on physicians. Notably, a recent study by Linder et al. found that at the end of a long shift, doctors are less able to make appropriate medical decisions: a phenomenon known as *decision fatigue* (2). Therefore, some contend that the long hours and overwhelming workload placed upon medical professionals impedes their ability to provide safe and proper healthcare (3).This is particularly concerning in the operating room, where medical staff must maintain acute motor and concentration skills for prolonged durations of time—and where mistakes can be deadly (4). As tired doctors may potentially put their patients at increased, unnecessary risk, we wondered whether surgeries performed later in the day were associated with increased mortality, as doctors are likely to be more tired at the end of a long day at work. Understanding the impact of physician fatigue on clinical outcomes may inform hospital policy change and protect patients from unnecessary medical mistakes.

To investigate the relationship between time of day and surgery mortality, we used data from a study by Sessler et. al (1) that included 32,001 elective surgeries performed at the Cleveland Clinic between January 2005 and September 2010. The data recorded the time of day that the surgery was recorded and whether patients died within 30 days of surgery. Patient characteristics such as BMI, Charlson Comorbidity Index (a health classification that rates a patient's relative fitness before surgery based on comorbid conditions), RSI (risk stratification index, or a patient's relative risk of dying during surgery), age, and gender were also collected.

The study was conducted on surgeries performed during the typical hospital workday between 6 AM and 7 PM. Though time was originally recorded as a continuous variable, we stratified time of day for our formal analysis: early day (before 11 AM), midday: between (11 AM and 3 PM), and late day (between 3 PM and 7 PM). We created these distinctions so that our results would be more clinically relevant, as a difference in mortality minute to minute would likely not be as applicable as a difference between surgeries performed in the morning versus the evening, and we additionally sought to create three relatively equal time periods.

We began with exploratory data analysis and visualization, noting that surgeries that resulted in mortality generally occurred later in the day (Figures 1 and 2). Based on our background research and exploratory analysis, we hypothesized that surgeries performed at the end of the workday would have a higher mortality rate than surgeries performed early in the day, even when controlling for any confounding variables (such as a patient's inherent risk of dying during the surgery).

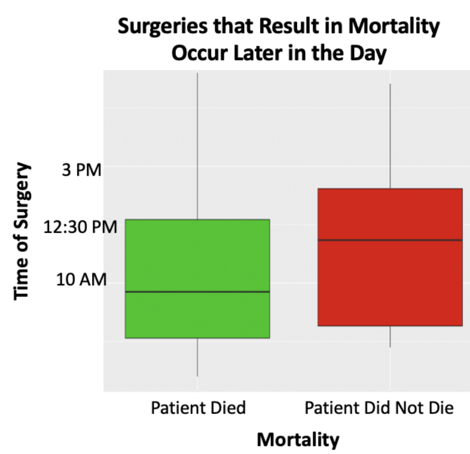

Figure 1: Plot of the time of day a patient received surgery vs. whether they died within 30 days of surgery

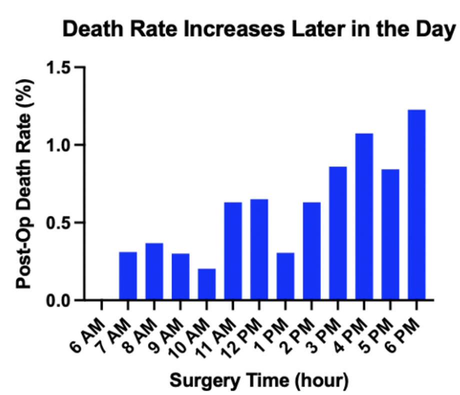

Figure 2: Graph of the percentage of patients who died within 30 days of surgery based on the hour in which they received surgery

II. Statistical Methodology

All analyses were performed at a significance level of 0.05. We first evaluate whether a general difference in mortality rate based upon the time of day a surgery was completed using a chi-square test. If sufficient evidence is found for differential mortality, we will proceed to then control for potential confounding variables using a logistic regression model that evaluates the relationship between the expected logit of the probability that a surgery resulted in mortality and the time of day in which it was performed, controlling for patient pre-op mortality risk score (RSI), Charlson Comorbidity Index (CCI), BMI, and age. We chose the additional predictors of interest as we thought they were the most important to a patient's inherent risk of dying during surgery in our model. Because we hypothesized increased mortality being associated with later surgeries as surgeons got tired at the end of the day *even after* controlling for potential confounders, we expected to see a significant difference in the log-odds of mortality between our three categories.

III. Results

Our initial chi-square test found sufficient evidence to suggest an association between time of day and 30 day mortality ($\chi^2_{2\ d.f.} = 27.3$; $p < 0.001$) and so we proceeded to fit the logistic regression model. Controlling for perioperative patient characteristics, patients having midday surgeries were expected to have 1.591 times the odds (95% CI: 1.179, 2.003) of mortality compared to patients having early day surgeries, and patients having late day surgeries were expected to have 2.848 times the odds (95% CI: 1.764, 2.848) of mortality compared to patients having early day surgeries. Unsurprisingly, higher pre-op risk scores, older age and higher Charlson Comorbidity Index scores were also associated with higher conditional log-odds of death. Full model output for the logistic regression model is available in the Appendix.

IV. Discussion

Based on our model, we found sufficient evidence to suggest a difference in the expected log-odds of mortality for patients who had surgery in the middle of the day compared to patients who had surgery in the morning, controlling for risk score, Charlson Comorbidity Index, BMI and age. We also found sufficient evidence to suggest a difference in the expected log-odds of mortality for patients who had surgery at the end of the day compared to patients who had surgery in the morning, while controlling for the same confounders.

In comparing the absolute magnitude of identified differences, midday surgeries were associated with an approximate 1.5-fold increase in expected mortality odds compared to early day surgeries and late day surgeries were associated with approximately double the expected odds of mortality. Though we may have found statistical significance in both the early vs. mid and early vs. late day comparisons, our findings may not be clinically relevant on a scale of absolute deaths given the relatively low probability of death in these surgeries (Figure 1); even surgeries performed late in the day have low mortality, with mortality (estimated hourly) never being over 1.5%.

Despite the observational nature of our data, we hypothesize that physician fatigue might contribute to the increased observed mortality, as physicians are more likely to make mistakes when they are tired. Unfortunately, it would be unethical to conduct a randomized study assigning "tiredness" states to physicians. Regardless, in the future, it may be important to explore the potential positive impact of decreasing physician shift length or mandating frequent breaks for physicians throughout the workday. It may also be interesting to explore the impact of decision fatigue (3) on this increase in mortality. In the meantime, it may also be prudent to schedule more surgeries earlier in the day, before physicians get fatigued.

In our analysis, we identified limitations stemming from the original study design and our own methods. First, the study itself only focused on elective surgeries performed during the day. We could have seen more significant or more widely applicable results if the data included riskier, non-elective surgeries or emergency surgeries that occurred during the night shift. Second, the study does not include information on shift timing or when people take their lunch break, as that could help to mitigate fatigue. We assumed that physicians work the entire shift, from 6 AM to 7 PM; however, this may not be true. Accounting for breaks and shift changes could have made our analysis more relevant. In Linder et al.'s study on decision fatigue, for example, a lunch break helped mitigate the negative effects of decision fatigue (3). Thus, in the future, a study that measured mortality risk before and after lunch breaks or shift changes, for example, might be important.

Although we split up the day into three dummy variables in order to simplify our analysis and make our results more applicable, this simplification comes with inherent limitations. By grouping time of day into three distinct variables instead of leaving time as a continuous variable or in smaller increments, we may have missed changes in mortality that occur from hour to hour, for example. In addition, in order to avoid multicollinearity, we only controlled for a subset of variables available, but could also have controlled for more variables such as ASA status, which may have been a confounding variable that we did not include.

Regardless, in our analysis we identified statistically significant increased odds of mortality for surgeries performed both in the middle of the day and later in the day when compared to surgeries performed early in the morning, while adjusting for patient and perioperative characteristics. These findings may motivate future hospital policy changes, such as a restructuring of surgery scheduling to earlier in the day or increased breaks, and may also inform future research on methods to mitigate physician fatigue.

References

1. Sessler, Daniel I. MD et al., Operation Timing and 30-Day Mortality After Elective General Surgery, Anesthesia & Analgesia: December 2011 - Volume 113 - Issue 6 - p 1423-1428 doi: 10.1213/ANE.0b013e3182315a6d
2. Mitler MM et al.. Catastrophes, sleep, and public policy: consensus report. Sleep. 1988 Feb;11(1):100-9. doi: 10.1093/sleep/11.1.100. PMID: 3283909; PMCID: PMC2517096.
3. Linder, J. A., Doctor, J. N., Friedberg, M. W., Reyes Nieva, H., Birks, C., Meeker, D., & Fox, C. R. (2014). Time of day and the decision to prescribe antibiotics. JAMA internal medicine, 174(12), 2029–2031. https://doi.org/10.1001/jamainternmed.2014.5225
4. Gaba DM, Howard SK. Patient safety: fatigue among clinicians and the safety of patients. N Engl J Med. 2002 Oct 17;347(16):1249-55. doi: 10.1056/NEJMsa020846. PMID: 12393823.
5. Bruce M Lo, M. D. (2021, October 16). Deep venous thrombosis risk stratification: Risk stratification. Deep Venous Thrombosis Risk Stratification: Risk Stratification. Retrieved April 25, 2022, from https://emedicine.medscape.com/article/1918446-overview

Appendix

APPENDIX: Logistic Regression Model

Independence is reasonable, as one patient's outcomes are unlikely to be related to another patient's outcomes. We assume a linear relationship between the predictors and the log-odds of mortality. Full model coefficients are as follows:

	Estimate	Standard Error	p-value
Intercept	-5.457	0.765	<0.001
Time of day: Early	Reference		
Time of day: Mid	0.467	0.210	0.027
Time of day: Late	0.835	0.276	0.003
Risk Score: High	Reference		
Risk Score: Med	-4.896	0.374	<0.001
Risk Score: Low	-2.707	0.304	<0.001
BMI (kg/m^2)	0.007	0.013	0.604
CCI	0.113	0.033	<0.001
Age (Years)	0.046	0.008	<0.001

Note: the Risk Score corresponds to a patient's pre-operation RSI score. In our dataset, patient risk score was calculated based on the International Classification of Diseases (9th rev.) codes. In our study, we stratified these risk scores into three distinct variables for simplicity and clinical relevance. We based this stratification on a similar study, which used an index with a similar methodology and range (5). In this study, patients with a RSI of 0 or lower were considered low-risk, patients with a RSI between 1 and 2 were medium-risk, and 3 or greater were considered high risk. In order to make sure we were accounting for all patients, we extended medium risk to include anyone with a RSI greater than 0 or less than 3. Thus, the following RSI scores were used for our study:

1. Low Risk: RSI score of less than 0
2. Medium Risk: RSI score between 0 and 3
3. High Risk: RSI score of 3 or greater