Online Supplement: Adaptive Behaviour of Service Providers to Schedule Deviations and Its Consequences: Evidence from Operating Rooms

Yiwen Jin, Yichuan Ding, Steven M. Shechter, Jugpal S. Arneja

EC.1. Operating Room Scheduling Practices

The shift's start time, marked by the commencement of the first surgery, is typically scheduled for either 7:45 am or 8:20 am. Occasionally, shifts may begin later due to additional training or administrative tasks. Prior to the shift's start, the surgical team receives a comprehensive briefing on the day's surgical schedule, even though the schedule is usually sent to the team several days in advance. The shift is usually scheduled to end between 3:15 pm - 3:45 pm. Patients with operations scheduled for that day will be notified a few days earlier and typically arrive about two to three hours before their scheduled surgery start times.

Scheduling surgeries in a hospital is a complex and lengthy process managed by a centralized booking office. There are nine ORs for regular shifts and two additional ORs for emergent add-on cases. Usually, each OR is assigned to a sole surgical team from one department throughout the day (shift). Each shift typically includes four or more surgeries. The booking office then contacts the patient and the surgeon to determine which surgeries will be performed on that date , ensuring the total shift length close to eight hours.

The booking office then drafts an *OR slate* for each surgery day in an online system. A surgery slate is a spreadsheet containing essential information for each surgery case listed in each row. This information includes planned start and end times, planned procedure duration, and a brief procedure description. It also contains other identifications such as patient name, a medical record number (MRN in Figure EC.1), bed number, and the surgeon in charge. Figure EC.1 presents an example of an OR slate with three surgeries scheduled for January 12, 2022.

We next introduce the current scheduling practice, which includes key decisions such as determining the job allowance for each surgery, selecting an appropriate number of surgeries to be performed in a shift from the waitlist, and sequencing them. As verified by the data, the job allowance is the summation of the estimated surgical duration and estimated turnover duration. The former is calculated using the median duration of all procedures of the same type that take place in the studied hospital in the previous calendar year; while the latter is set to either 20, 25, or 30 minutes, depending on the procedure type. The booking office and the surgeon jointly select patients from the waitlist to perform surgeries and such selection is usually made by the wait-to-target ratio according to the internal hospital document. Different patients have different wait time targets based on their urgency levels. The wait-to-target ratio then serves as a way for the surgeons to normalize offsite wait times relative to their target wait time, which affects how they prioritize patients. An appropriate number of surgeries will be booked so that the total job allowance can fit roughly into an 8-hour shift. We do not have complete information on the sequencing of surgeries within a shift. However, our empirical analysis based on the data indicates no clear difference in terms of job allowances, running early or late, nor reoperation rates across the sequences within a shift as shown in Table EC.3 and Table EC.4. (The increasing trend of the coefficients in Table EC.3, though not statistically significant, is attributed to the rising standard deviation as the sample size decreases with higher surgery indices.)

OR SLATE

From 12-Jan-2022 to 13-Jan-2022

Total # Cases: 3

Case #	Surgeons	Start Time	End Time	PIR (min)	Patient Name	MRN	Procedure	Bed
January 12, 2022								
1234567	Surgeon A	7:45	10:00	135	DOE, JOHN	15645611	P1	20-ADP-PICU
9876541	Surgeon A	10:30	12:30	120	SMITH, JACK	25654894	P1	10-SDC
9852642	Surgeon A	13:00	15:00	120	TURNER, MARY	34948480	P3	30-INPT

Figure EC.1 An Example of an OR Slate

The completed day-shift schedule is distributed to the surgical and cleaning teams at least one day before the surgery day so that they are well-informed about each surgery's planned start and end time. The finalized schedules are then recorded in the system and appear in the dataset. Occasionally, some emergency cases are added to the shift on the surgery day without notifying the surgical team in advance. These emergent add-on cases are not recorded on the slate since they have no advanced schedule. However, they are still recorded in the surgical database. About 7.6% of the shifts contain add-on cases. We discuss how to handle shifts with add-on cases in Section 4.2.

EC.2. GMM Estimation

For dynamic panel data, the classic fixed-effect ordinary-least-square (FE-OLS) estimator yields biased estimators, especially when the panel data has large cross-sectional units N and short time periods T (in our case, $N = 1,928$ and $T \in [3,8]$). This bias is known as the *Nickell bias* (Nickell 1981). The bias arises because the lagged response variable is one of the explanatory variables. Therefore after the demeaning process (a step of the FE estimator), there will be a correlation between the autoregressive regressor and the error term. Thus, to achieve unbiased estimates of γ_n and γ_p from Equation (EC.2), we adopt the Arellano-Bond (A-B) method (Arellano and Bond 1991) of dynamic panel models. The A-B approach utilizes higher-order lagged regressors as instrumental variables (IV) that induce a series of moment conditions for GMM estimation. The IV selection criterion depends on the properties of the explanatory variables. We now take the model of RPD as an example to elaborate the GMM estimation process. For RTD, the estimation process is largely the same. First, we transform the original model Equation (EC.2) via first-difference (FD) or forward-orthogonal deviation (FOD). We try both approaches to reach a robust estimate. FD is a simple and direct method by taking the first order difference between two consecutive cases, which is also applied to regular static panel models (Anderson and Hsiao 1981, Arellano and Bond 1991).

Arellano and Bover (1995) proposed FOD method for dynamic panel data and the estimator is called the FOD-GMM estimator. The FOD transformation operator is as follows

$$\tilde{\Delta}_t u_{it} = \sqrt{\frac{T-t+1}{T-t}} \left(u_{it} - \frac{1}{T-t+1} \sum_{s=0}^{T-t} u_{i,t+s} \right)$$

Under the assumption that u_{it} is serially uncorrelated, we obtain that $\text{Corr}\left(\tilde{\Delta}_t u_{it}, \tilde{\Delta}_t u_{i,t-1}\right) = 0$. The advantage that the FOD-GMM estimator has compared with diff-GMM is the higher data utilization rate for unbalanced panel data sets. We use S_{it} to denote $DSS_{i,t+1}$ and we set S_{i0}, the DSS of the first case in a shift as the difference between the real start time and scheduled start time so that more data can be utilized.

As $\text{DSS}_{i,t}$ is a linear function of $\text{RPD}_{i0}, \text{RPD}_{i1}, \ldots, \text{RPD}_{i,t-1}$ as given in Equation (1), multiple lagged response variables $\text{RPD}_{ik}, k = 1, \ldots, t-1$ appear as explanatory variables of the regression. In this case, the OLS method leads

to estimation bias (Nickell 1981). To address this issue, we reformulate the problem as an auto-regressive specification using a linear transformation. From Equation (1), we have

$$\mathrm{DSS}_{i,t+1} - \mathrm{RTD}_{it} = \mathrm{DSS}_{i,t} + \mathrm{RPD}_{it}, \quad t = 1, 2, \ldots, T_i. \tag{EC.1}$$

Plugging the expression of RTD_{it} in Equation (2) into the right hand side, we obtain

$$\mathrm{DSS}_{i,t+1} - \mathrm{RTD}_{it} = (\gamma_n - 1) \cdot (\mathrm{DSS}_{i,t})^- + (\gamma_p + 1) \cdot (\mathrm{DSS}_{i,t})^+ + \mathbf{x}_{it}^\top \beta + \delta_t + \alpha_i + u_{it}, \tag{EC.2}$$

which is an auto-regressive model often employed in dynamic panels (Pesaran 2015).

For the relative turnover duration (RTD), with a slight abuse of notation, we have the following similar specification:

$$\mathrm{RTD}_{it} = \nu_n \cdot (\mathrm{DSE}_{i,t-1})^- + \nu_p \cdot (\mathrm{DSE}_{i,t-1})^+ + \mathbf{x}_{it}^\top \beta_1 + \mathbf{n}_{it}^\top \beta_2 + \tilde{\delta}_t + \tilde{\alpha}_i + \tilde{u}_{it}, \ t = 1, \ldots, T_i, \tag{EC.3}$$

where DSE stands for deviation from the scheduled exit time, which is exactly the scheduled start time of the turnover stage. Apart from the same regressors $\mathbf{x}_{it}$ as in Equation (2), we also include $\mathbf{n}_{it}$ for RTD that comprises 1) indicator of whether case t and case $t+1$ belong to the same type and 2) the patient type of case $i+1$. This is because the turnover stage consists of both cleaning and preparation steps for two consecutive operations. Like DSS, a positive DSE means that the focal surgery finishes late (thus the subsequent turnover stage starts late) and vice versa. The coefficients of interest are ν_n and ν_p for the negative and positive parts, respectively. The relationship between DSE and RTD is $\mathrm{DSE}_{i,t} = \mathrm{DSE}_{i,t-1} + \mathrm{RTD}_{i,t} + \mathrm{RPD}_{i,t+1}$. After similar transformation processes, we obtain the following auto-regressive model for RTD:

$$\begin{aligned} \mathrm{DSE}_{i,t} - \mathrm{RPD}_{i,t+1} = (\nu_n - 1) \cdot (\mathrm{DSE}_{i,t-1})^- + (\nu_p + 1) \cdot (\mathrm{DSE}_{i,t-1})^+ + \\ \mathbf{x}_{it}^\top \beta_1 + \mathbf{n}_{it}^\top \beta_2 + \tilde{\delta}_t + \tilde{\alpha}_i + \tilde{u}_{it}, t = 1, \ldots, T_i \end{aligned} \tag{EC.4}$$

To obtain unbiased estimates of γ_n, γ_p in Equation (EC.2) and ν_n, ν_p in Equation (EC.4), we employ the Arellano-Bond (A-B) generalized method of moments (GMM) for dynamic panel data (Arellano and Bond 1991). Essentially, the A-B method utilizes higher-order lagged regressors as instrumental variables (IVs) that induce a series of moment conditions for GMM estimation. It is important to note that the directly obtained estimates are $\hat{\gamma}_n - 1$ and $\hat{\gamma}_p + 1$ for RPD ($\hat{\nu}_n - 1$ and $\hat{\nu}_p + 1$ for RTD), while we report $\hat{\gamma}_n, \hat{\gamma}_p, \hat{\nu}_n, \hat{\nu}_p$ after adding (subtracting) one.

After transformation, we obtain the following two equations for them, respectively,

$$\begin{aligned} \mathrm{FD}: \ \Delta(\mathrm{DSS}_{i,t+1} - \mathrm{RTD}_{it}) &= (\gamma_n - 1)\Delta(\mathrm{DSS}_{i,t})^- + (\gamma_p + 1)\Delta(\mathrm{DSS}_{i,t})^+ + \Delta\mathbf{x}_{it}^\top \beta + \Delta\delta_t + \Delta u_{it} \\ \mathrm{FOD}: \ \tilde{\Delta}_t(\mathrm{DSS}_{i,t+1} - \mathrm{RTD}_{it}) &= (\gamma_n - 1)\tilde{\Delta}_t(\mathrm{DSS}_{i,t})^- + (\gamma_p + 1)\tilde{\Delta}_t(\mathrm{DSS}_{i,t})^+ + \tilde{\Delta}_t\mathbf{x}_{it}^\top \beta + \tilde{\Delta}_t\delta_t + \tilde{\Delta}_t u_{it}, \end{aligned} \tag{EC.5}$$

where Δ and $\tilde{\Delta}_t$ stand for the FD and FOD transformation operators, respectively. Both FD and FOD transformations eliminate all time-invariant fixed effects such as α_i in Equation (2). We control for other variables that vary with case index (time) t in the same shift, such as patient type, case type, and case index, by including them in $\mathbf{x}_{it}$, which remain in the regression equations after the FD or FOD transformation.

Second, we need to determine the appropriate lag order of all explanatory variables on the RHS of Equation (EC.5), which leads to the moment conditions that are indispensable for GMM estimation in the A-B method. Generally speaking, the right-hand-side regressors $y_{i,t-1}$ and $\mathbf{x}_{it}$ can be categorized into four types further for which the moment conditions are different. (i). *Lagged dependent variable*: $E[y_{i,t-s}\Delta u_{it}] = 0, \quad s = 2, 3, \ldots, t$; (ii). *Strictly exogenous regressors*: $E[\mathbf{x}_{i,t-s}\Delta u_{it}] = \mathbf{0}, \quad t - s = 1, \ldots, T$; (iii). *Predetermined regressors*: $E[\mathbf{x}_{i,t-s}\Delta u_{it}] = \mathbf{0}, \quad s = 1, 2, \ldots, t$; (iv). *Endogenous regressors*: $E[\mathbf{x}_{i,t-s}\Delta u_{it}] = \mathbf{0}, \quad s = 2, 3, \ldots, t$. The criterion that determines the category of a regressor is condition independence where strictly exogenous regressors satisfy $E[u_{it} \mid \mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}] = 0$, that is,

all of them are conditionally independent of the current error term at t. Predetermined regressors should satisfy $E[u_{it} \mid \mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}] = 0$ and endogenous ones satisfy $E[u_{it} \mid \mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i,t-1}] = 0$.

Here we consider higher lags of the auto-regressive term, namely $(S_{i,t-2})^-, (S_{i,t-3})^-$ and $(S_{i,t-2})^+, (S_{i,t-3})^+$, as instruments for $\Delta(S_{i,t-1})^-$ and $\Delta(S_{i,t-1})^+$ in Equation (EC.5). The associated moment conditions are formulated in Equation (EC.6). Regressors in $\mathbf{x}_{it}$ include exogenous variables of case t such as the patient age, surgery procedure, and case types, which are uncorrelated with u_{it}. We select these explanatory variables and first-order lagged terms as instruments and formulate their moment conditions. The moment conditions for the FD transformed model are presented in Equation (EC.6) below:

$$E\left[(S_{i,t-s})^- \Delta u_{it}\right] = 0, E\left[(S_{i,t-s})^+ \Delta u_{it}\right] = 0, s = 2,3; E\left[\mathbf{x}_{i,t-s} \Delta u_{it}\right] = 0, s = 1,2 \tag{EC.6}$$

In the FOD model, by subtracting forward means from the current levels, we can utilize the first-order lagged terms as instruments, and the associated moment conditions are as follows:

$$E\left[(S_{i,t-s})^- \tilde{\Delta}_t u_{it}\right] = 0, E\left[(S_{i,t-s})^+ \tilde{\Delta}_t u_{it}\right] = 0, s = 1,2; E\left[\mathbf{x}_{i,t-s} \tilde{\Delta}_t u_{it}\right] = 0, s = 0,1 \tag{EC.7}$$

In fact, we can adopt further lagged terms as instruments. Nevertheless, too many instruments may lead to over-identification issues in estimation. Thus, for endogenous variables, we only use their lags of the second and third order for the FD-GMM model, and we use lags of the first and second order for the FOD-GMM model. While for exogenous variables, we always use themselves and first-order lagged terms as instruments. We conduct a series of specification tests to validate the choice of lagged terms (see Section EC.3).

We then conduct GMM estimation based on the above moment conditions. After stacking the moment conditions, we obtain the vector-form moment conditions for GMM estimation:

$$E[\mathbf{m}_i(\boldsymbol{\theta})] = \mathbf{0}, \quad \boldsymbol{\theta} = (\gamma, \boldsymbol{\beta})$$

where $\boldsymbol{\theta}$ is the vector of all estimands in Equation (EC.5), $\mathbf{m}_i(\cdot), i = 1,2,\ldots,N, N = 7,868$ denotes the left-hand-side of the moment conditions (EC.6) applied to a case i in the FD-GMM model (or Equation (EC.7) in the FOD-GMM model). The GMM estimators are obtained through the following two-step minimization:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\mathbf{b}} \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_i(\mathbf{b})\right)^\top \mathbf{W}_0 \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_i(\mathbf{b})\right)$$

$$\hat{\boldsymbol{\theta}}_{step2} = \arg\min_{\mathbf{b}} \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_i(\mathbf{b})\right)^\top \mathbf{W}(\hat{\boldsymbol{\theta}}) \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_i(\mathbf{b})\right), \mathbf{W}(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_i(\hat{\boldsymbol{\theta}})\mathbf{m}_i(\hat{\boldsymbol{\theta}})^\top\right)^{-1}.$$

In the first step of the GMM, we utilize an initial (suboptimal) weighting matrix $\mathbf{W}_0$, and the purpose is to obtain a GMM estimate of $\boldsymbol{\theta}$. We then obtain the consistent estimate of the inverse of the asymptotic covariance matrix of $\mathbf{m}(\hat{\boldsymbol{\theta}})$, as shown in the second set of equations. We finally update $\mathbf{W}_0$ by $\mathbf{W}(\hat{\boldsymbol{\theta}})$ and obtain the two-step GMM estimator. More details regarding GMM computation can be found inArellano and Bond (1991), Arellano and Bover (1995), Windmeijer (2005), Kripfganz et al. (2019).

EC.2.1. Heterogeneity Across Surgeon Seniority

We explore the responsive behavior across different surgeon seniority groups. The surgeon is the leader of the surgical team during an operation. Thus, the surgeons' experience is vital to guide other members. In this analysis, we focus on the RPD. There is extensive literature on the relationship between experience and performance, especially in healthcare settings (Kc and Staats 2012, Staats et al. 2018, Ibanez et al. 2018). Senior surgeons with more experience

are more familiar with the procedures. Therefore, they are conscious of both the difficulties and the more manageable parts of a procedure.

To investigate how surgical teams' behavior varies by surgeon seniority, we perform a subsample study for junior and senior surgeon groups where we choose 20 years after obtaining the doctor of medicine (M.D.) degree as the threshold of surgeon seniority. In our setting, the junior (< 20) and senior (≥ 20) groups have 3,793 and 4,075 surgical cases, respectively.

Table EC.1 Heterogeneity of Responses in Surgery Stage between Surgeon Seniority Groups

	Senior		Junior	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.679***	0.505***	0.368	0.404
	(0.150)	(0.141)	(0.265)	(0.210)
Delayed Start	-0.359*	-0.319**	-0.313*	-0.234
	(0.155)	(0.113)	(0.154)	(0.136)
Elective case	59.10*	42.09	15.79	-13.77
	(23.89)	(37.86)	(65.03)	(48.12)
Sequence shuffling	-5.889	-4.485	0.166	-3.363
	(7.337)	(5.698)	(10.28)	(8.709)
Age	-0.255	-0.129	0.324	0.410
	(0.273)	(0.189)	(0.247)	(0.224)
Number of instruments	244	244	227	227
Number of observations	3,793	3,793	4,075	4,075

Notes: Patient type, patient severity, procedure type, and time fixed effect have been included across all specifications. Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

From Table EC.1, we find that the senior surgeons' group yields significant estimates for both early and delayed start. To understand the scale of such adaptive behavior, again, we consider one SD delayed (early) start as an example. Senior surgeons speed up by $0.319 \times 23.03 = 7.35$ minutes (10.97% of median procedure duration) or slow down by $0.505 \times 15.88 = 8.02$ minutes (11.97% of median procedure duration) in the focal case on average. Surgical teams led by senior surgeons tend to expedite when facing delays and slow down when getting ahead of schedule at a similar pace. On the other hand, junior surgeons do not always present speedup or slowdown behavior. As pointed out by our surgeon collaborators, an explanation of such discrepancies between their responsive patterns is that compared to junior surgeons, senior surgeons may have trainees in the OR. Senior surgeons would be faster and likely allow trainees to participate in the procedures for better practice. When they are running late, the senior surgeon may take over from the trainee to navigate the schedule back on track. Junior surgeons, nevertheless, might be doing the work themselves without substantially involving the trainees.

EC.3. Robustness Checks

We further consider whether the hospital strategically plans surgeries by assigning the sequence of surgeries based on complexity in one day's shift. For example, complex surgeries (more likely to run overtime) are followed by the more straightforward surgeries (more likely to finish early). As shown in Table EC.3, we do not find there exist significant trends of relative procedure duration (RPD) or the planned job allowance across different sequences.

We also examine if the reoperation rate significantly varies across sequences using ANOVA. We find that there is no significant differences between subgroup averages as shown in Table EC.4.

Table EC.2 Heterogeneity and Nonlinearity of Responsive Behavior

	Heterogeneity by EOS		Nonlinearity	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.875***	0.439**	0.510**	0.425***
	(0.219)	(0.163)	(0.148)	(0.120)
Early Start × Late Shift	-0.852*	-0.644*		
(planned EOS after 4 pm)	(0.406)	(0.318)		
Delayed Start	-0.298**	-0.189*		
	(0.115)	(0.096)		
Delayed Start × Late Shift	-0.023	-0.190		
(planned EOS after 4 pm)	(0.409)	(0.297)		
Delayed Start Spline:				
0-18 mins			-0.275*	-0.244*
			(0.137)	(0.123)
≥ 18 mins			-0.255*	-0.176*
			(0.100)	(0.815)
Number of instruments	259	255	261	257
Number of observations	7,868	7,868	7,868	7,868

Notes: Columns (1)-(2) report results for heterogeneity by planned EOS; Columns (3)-(4) report results for nonlinearity. All specifications include controls for: elective case status, sequence shuffling, patient age, patient type, patient severity, procedure type, and time fixed effects. Both GMM models are estimated using the Arellano-Bond method. Time fixed effect t stands for the sequence of cases in shift i. Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

Table EC.3 Relationship between Sequence, RPD and Job Allowance

Variables	RPD	Job Allowance
Sequence: 2nd	-1.5315	-3.8226
	(1.9631)	(2.2906)
Sequence: 3rd	-0.1110	-1.3826
	(2.3703)	(2.9229)
Sequence: 4th	0.5563	-6.1770
	(2.6691)	(3.2058)
Sequence: 5th	2.5842	-6.8662*
	(2.8445)	(3.3159)
Sequence: 6th	2.7723	-4.7536
	(3.0562)	(3.3813)
Sequence: 7th	2.6408	-4.3722
	(3.1684)	(3.7337)
Sequence: 8th	3.5412	-4.9592
	(3.9971)	(4.8159)
R^2	0.1195	0.6949
N	7,866	7,866

Notes: Robust standard errors clustered by shift are shown in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). Both models include controls for: elective case, sequence shuffling, age, patient type, patient severity, procedure type, and hour-of-day.

EC.3.1. Dynamic Panel Model Specification Tests

Serial Correlation Test. Recall that we utilize higher-order lagged regressors as instruments in Equation (EC.6) for GMM estimation of the dynamic panel model. Under such an IV configuration, for these lagged explanatory variables to be valid instruments, the absence of higher-order serial correlation between error terms Δu_{it} and $\Delta u_{i,t-s}, (s=2,3)$

Table EC.4 Reoperation Rates across Sequences

	Sequence					
Statistic	1	2	3	4	5	6
Average	0.0369	0.0240	0.0349	0.0225	0.0310	0.0099
Std. dev.	0.1886	0.1532	0.1835	0.1484	0.1736	0.0993
F-statistics: 1.61, p-value: 0.1268						

Table EC.5 FE-OLS Estimation Results

Variables	Surgery	Turnover
Early start	0.430***	0.283***
	(0.035)	(0.018)
Delayed start	-0.533***	-0.059***
	(0.023)	(0.013)
Elective case	13.49	6.419*
	(10.25)	(2.948)
Sequence shuffling	-2.353	-5.208**
	(2.217)	(1.589)
Age	0.348***	0.060
	(0.091)	(0.046)
Different procedure		1.175**
		(0.435)
R^2	0.496	0.531
N	7,787	6,120

Notes: The linear model Equation (2) and Equation (EC.3) are estimated by OLS. Estimated by FE-OLS. We also control for: patient type, patient severity, procedure type, shift and time fixed effects. "Next case type" is included only in the Turnover specification. Robust standard errors clustered by shift are shown in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

is indispensable. Here the formal null hypothesis is

$$H_0 : \mathrm{Corr}\left(\Delta u_{it}, \Delta u_{i,t-j}\right) = 0,\ j = 2, 3. \tag{EC.8}$$

We thus adopt the Arellano-Bond specification test, also known as the AR(p) test, to examine the above validity requirements, where p stands for the lagged order. To be specific, we perform AR(1) to AR(3) tests for both Models (1) and (2) to identify the starting level of the lag structure of the instruments. From the p values in the last part of Table EC.6 we find that the AR(1) test is rejected, since applying FD transformation to Equation (2) naturally brings first order serial correlation between Δu_{it} and $\Delta u_{i,t-1}$. We cannot reject the null hypotheses of AR(2) and AR(3), which means there is no higher order serial correlation in Equation (EC.5). Therefore, our null hypothesis Equation (EC.8) is supported and the IV configuration we choose is valid for identification. For moment equations of the FOD-GMM model we also conduct similar tests and the results also cannot reject the null hypothesis.

We also conduct the Hansen J-test (Hansen 1982) of model over-identification, as the number of instruments is relatively large. The null hypothesis is H_0 : Over-identifying restrictions are valid. We see that both the FD-GMM and FOD-GMM models yield non-significant χ^2 values. Thus they pass the Hansen test and the IV configurations for identification are supported.

EC.3.2. Sample Inclusion Criterion

Remove all shifts with add-on cases. We remove the shifts that include add-on cases from our sample, which yields 7,484 cases (5,827 for RTD) in all. The results are shown in Table EC.7. We find that the estimates for both surgery

Table EC.6 Specification Tests for Main Model

	Surgery		Turnover	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Specification tests:				
Arellano–Bond AR(1) p value	0.000	0.000	0.000	0.000
Arellano–Bond AR(2) p value	0.593	0.299	0.205	0.589
Arellano–Bond AR(3) p value	0.950	0.868	0.951	0.436
Hansen J-test χ^2	129.2	124.6	135.2	108.6
Hansen J-test p value	0.287	0.392	0.163	0.649
Number of instruments	253	253	258	253
Number of observations	7,868	7,868	6,210	6,210

stage (RPD) and turnover stage (RTD) are consistent with Table 2 and significant. Both surgical and cleaning teams respond to both directions.

Remove all shifts affected by cancellations. We remove all shifts that are affected by cancellations, which yields 7,777 cases in the refined sample (6,122 for RTD). The GMM estimation results are shown in Table EC.7. We find that the estimates are still consistent. These results support our findings of the adaptive behavior by surgical and cleaning teams.

EC.3.3. Parallel ORs

We consider competing demands for shared resources across ORs to further test the robustness of our results. Specifically, we define parallel ORs as follows: during the procedure stage, it is the number of ORs whose shifts are concurrently ongoing. For the turnover stage, it equals the number of other ORs whose turnover windows overlap the focal turnover interval—i.e., those still in turnover at the focal turnover's start or beginning turnover before it ends. This approach approximates the number of parallel ORs competing for shared resources. The estimation results are shown in Table EC.8. The speed-up/slow-down effects remain unchanged when this control is included, and the coefficient of the number of parallel ORs is statistically insignificant across specifications.

EC.3.4. IV Specification Tests

To verify the two IVs: DSS and the waitlist size when focal patient joins, are not correlated with the unobserved confounders, we run a regression by controlling them and also the two key variables of interest in a plain probit model for 30-day reoperation. The results, as summarized in Table EC.9, show that the estimates of DSS and waitlist size are not significant. This means that the IVs are not correlated with other unobserved confounders on reoperations, which provides support for the validity of IVs.

EC.3.5. Alternative Reoperation Measures

To test the robustness of our results, we explore alternative time frames for reoperations. We use alternative post-surgical complication measures — 60-day and 90-day reoperations to examine the impact of surgical duration and wait times using the same specification in Equation (3). We find that our results are consistent under these alternative measures as shown in Table EC.10.

Table EC.7 Robustness Estimation Results

	Surgery		Turnover	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Panel A: Robustness with Add-on Cases				
Early Start	0.590***	0.424**	0.250***	0.306***
	(0.166)	(0.134)	(0.057)	(0.057)
Delayed Start	-0.242*	-0.167*	-0.120*	-0.088†
	(0.110)	(0.082)	(0.048)	(0.050)
Specification tests:				
Arellano–Bond AR(1) p value	0.000	0.000	0.000	0.000
Arellano–Bond AR(2) p value	0.627	0.367	0.430	0.514
Arellano–Bond AR(3) p value	0.812	0.610	0.858	0.488
Hansen J-test χ^2	129.3	121.0	123.5	119.9
Hansen J-test p value	0.286	0.482	0.394	0.435
Number of instruments	253	253	256	252
Number of observations	7,484	7,484	5,827	5,827
Panel B: Robustness with Cancellations				
Early Start	0.614***	0.499***	0.203***	0.291***
	(0.159)	(0.131)	(0.048)	(0.045)
Delayed Start	-0.212*	-0.170*	-0.098*	-0.081†
	(0.105)	(0.084)	(0.043)	(0.043)
Specification tests:				
Arellano–Bond AR(1) p value	0.000	0.000	0.000	0.000
Arellano–Bond AR(2) p value	0.630	0.551	0.194	0.581
Arellano–Bond AR(3) p value	0.977	0.608	0.806	0.519
Hansen J-test χ^2	128.2	120.3	133.8	108.5
Hansen J-test p value	0.311	0.500	0.184	0.653
Number of instruments	253	253	258	253
Number of observations	7,777	7,777	6,122	6,122

Notes: Panel A reports results for Add-on Cases; Panel B reports results for Cancellations. All specifications include controls for: elective case status, sequence shuffling, patient age, patient type, patient severity, procedure type, and time fixed effects. "Next case type" is included only in the Turnover specifications. Both GMM models are estimated using the Arellano-Bond method. Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses (†$p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$).

EC.4. Queueing Model Analysis

EC.4.1. Mathematical Formulation and Proofs

When the system is stable, following (Green and Savin 2008) and the Little's law, the number of jobs in steady state, denoted by $B(T)$, has the following distribution:

$$\pi(0,T) = \left[\sum_{l=0}^{\infty}\left(\frac{\lambda T}{(\prod_{i=0}^{l-1}(1-\Phi(\beta_0+\tilde{\beta}_1\min(i\cdot T,W)+\beta_2 T)))^{1/l}})\right)^{l}\right]^{-1}$$

$$\pi(k,T) = \pi(0,T)\prod_{i=0}^{k-1}\rho_i(T), \qquad \text{for } k\geq 1 \tag{EC.9}$$

where the effective utilization rate of the system at state i is

$$\rho_i(T) := \frac{\lambda T}{1-\Phi(\beta_0+\tilde{\beta}_1\min(i\cdot T,W)+\beta_2 T)} \tag{EC.10}$$

Table EC.8 GMM Estimation Results with the Number of Parallel ORs

	Surgery		Turnover	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.582***	0.440***	0.186***	0.288***
	(0.154)	(0.126)	(0.050)	(0.046)
Delayed Start	-0.260*	-0.174*	-0.107**	-0.098**
	(0.108)	(0.084)	(0.043)	(0.045)
Elective case	21.29	-13.96	4.604	-24.303
	(49.95)	(48.58)	(2.845)	(19.090)
Sequence shuffling	-2.601	-3.121	-7.130***	-9.480***
	(5.470)	(5.110)	(1.851)	(2.284)
Age	0.233	0.273	0.358**	-0.029
	(0.153)	(0.140)	(0.161)	(0.069)
Different procedure			0.967*	1.764***
			(0.585)	(0.524)
Number of parallel ORs	1.618	0.428	0.179	-0.103
	(1.149)	(0.940)	(0.262)	(0.564)
Number of instruments	255	255	260	257
Number of observations	7,868	7,868	6,210	6,210

Notes: Both GMM models are estimated using the Arellano-Bond method. In column (1) and (2), we report $\hat{\gamma}_n$ and $\hat{\gamma}_p$ for RPD. In column (3) and (4), we report $\hat{\nu}_n$ and $\hat{\nu}_p$ for RTD. All specifications include controls for: patient type, patient severity, procedure type, and time fixed effects. "Next case type" is included only in the Turnover specifications. Time fixed effect stands for the sequence of cases in shift i while the shift fixed effect is zeroed out in the FD (FOD) processes. Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

Table EC.9 IV Exclusion Test

	30-day Reoperation
Focal surgery procedure duration	1.7260***
	(0.3050)
DSS	0.0022
	(0.0013)
Relative turnover before focal surgery	0.0029
	(0.0045)
Wait-to-target	0.0015
	(0.0029)
Wait list size when joining (IV)	-0.0015
	(0.0030)
Sequence shuffling	Included
Age	Included
Patient type	Included
Patient severity	Included
Procedure type	Included
Time fixed effect	Included
Day-of-week fixed effect	Included
Surgeon fixed effect	Included
Number of observations	5,204

Notes: Estimated by the maximum likelihood method. Dependent variables: 30-day reoperation indicator. Robust standard errors are shown in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

With $\pi(k,T)$, we can compute the average number of jobs $E[B(T)]$ and expected response time $E[W(T)]$ as follows:

$$E[B(T)]=\sum_{k=0}^{\infty}k\cdot\pi(k,T)=\pi(0,T)\cdot\sum_{k=1}^{\infty}k\prod_{i=0}^{k-1}\rho_i(T)=\frac{\sum_{k=1}^{\infty}k\prod_{i=0}^{k-1}\rho_i(T)}{1+\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\rho_i(T)},$$

Table EC.10 Impact of Wait Times and Surgical Duration on Alternative Measures

Variables	60-day Reoperations	90-day Reoperations
Wait-to-target	-5.7974*	-5.9866**
	(2.8425)	(0.8996)
Procedure duration	0.0994**	0.1006**
	(0.0367)	(0.0207)
Sequence shuffling	Included	Included
Age	Included	Included
Patient type	Included	Included
Patient severity	Included	Included
Procedure type	Included	Included
Time fixed effect	Included	Included
Day-of-week fixed effect	Included	Included
Surgeon fixed effect	Included	Included
N	5,204	5,204

Notes: IV-probit model estimated by the maximum likelihood method. Robust standard errors are shown in parentheses (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

$$E[W(T)] = E[B(T)]/\lambda = \frac{\sum_{k=1}^{\infty} k \prod_{i=0}^{k-1} \rho_i(T)}{\lambda[1+\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\rho_i(T)]}, \tag{EC.11}$$

We further obtain the steady-state expected reoperation probabilities as follows.

$$\begin{aligned} E[R(T)] &= \sum_{k=0}^{\infty} r(k,T)\pi(k,T) \\ &= \pi(0,T)r(0,T) + \pi(0,T)\sum_{k=1}^{\infty} r(k,T)\left(\prod_{i=0}^{k-1}\rho_i(T)\right) \\ &= \frac{\Phi(\beta_0+\beta_2 T) + \sum_{k=1}^{\infty}\Phi(\beta_0+\tilde{\beta}_1\min(kT,W)+\beta_2 T)\prod_{i=0}^{k-1}\rho_i(T)}{1+\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\rho_i(T)}. \end{aligned} \tag{EC.12}$$

We present the proofs of the propositions in Section 6.2 as follows.

The following is the proof of Proposition 1.

Proof. We use t_l and t_u to represent the lower and upper bounds of varying T, which are the 20% decrease and 20% increase of the current value T_0. To achieve convergence of the expected backlog, the two infinite series $\sum_{k=1}^{\infty} k\prod_{i=0}^{k-1}\rho_i(T)$ and $\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\rho_i(T)$ in eq. (EC.11) must converge. We thus need

$$\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| < 1$$

where a_k stands for $k\prod_{i=0}^{k-1}\rho_i(T)$ or $\prod_{i=0}^{k-1}\rho_i(T)$ in the series.

We simplify the above ratio and obtain

$$r(T) := \lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = \frac{\lambda T}{1-\Phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T)} \geq \rho_i(T) \quad \forall i \tag{EC.13}$$

Note that for the numerator series we obtain the same ratio

$$\left|\frac{a_{k+1}}{a_k}\right| = \lambda T \lim_{k\to\infty}\frac{k+1}{k}\frac{1}{1-\Phi(\ldots)} = \frac{\lambda T}{1-\Phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T)}$$

Hence, to maintain convergence, we must have

$$\Phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T) < 1-\lambda T \tag{EC.14}$$

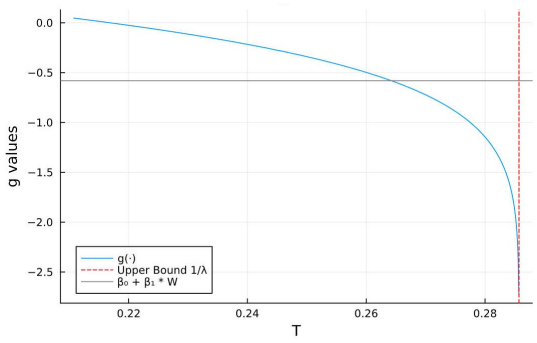

Figure EC.2 Function $g(T)$

$\Phi(\cdot) \in [0,1]$, $0 < 1-\lambda T < 1$ should hold to achieve convergence. Thus we cannot have very large service time ($T > 1/\lambda$) otherwise the system is not stable. When $T > 1/\lambda$ it's always diverging. Therefore we have $T \in [t_l, \min(t_u, 1/\lambda)]$.

Recall that we have denoted

$$g(T) := \Phi^{-1}(1-\lambda T) - \beta_2 T > \beta_0 + \tilde{\beta}_1 W \tag{EC.15}$$

Solving for its derivative we obtain

$$g'(t) = \frac{-\lambda}{\phi(\Phi^{-1}(1-\lambda t))} - \beta_2 = 0 \Rightarrow \Phi^{-1}(1-\lambda t) = \phi^{-1}(-\lambda/\beta_2)$$

Here, Φ^{-1} denotes finding the z-score corresponding to a PDF value, which generally doesn't have a closed-form solution and would typically be computed numerically. Also since $\phi \in (0, 1/\sqrt{2\pi}]$, when $-\lambda/\beta_2 < 1/\sqrt{2\pi}$, it is possible that $g(t)$ first increases then decreases after a tipping point (i.e., an inverted U-shape). Then we obtain two intersection points with $-\lambda/\beta_2$, which are denoted as t_{left} and t_{right}. The series eq. (6) is converging for $t \in [t_{left}, t_{right}] \cap [t_l, t_u]$.

When $-\lambda/\beta_2 \geq 1/\sqrt{2\pi}$, we have that $g(t)$ is monotonically decreasing. To achieve convergence, we require that $t \in [t_l, \min(g^{-1}(\beta_0 + \tilde{\beta}_1 W), t_u)]$.

At the maximum, we obtain

$$1 - \lambda t = \Phi(\phi^{-1}(-\lambda/\beta_2)) \Rightarrow t^* = \frac{1 - \Phi(\phi^{-1}(-\lambda/\beta_2))}{\lambda},$$

and the associated maximal value is $g(t^*) = \Phi^{-1}(1-\lambda t^*) - \beta_2 t^*$.

Therefore, when $-\lambda/\beta_2 \geq 1/\sqrt{2\pi}$, the convergence of expected backlog eq. (EC.11) requires that $\beta_0 + \tilde{\beta}_1 W < g(t_l)$. Therefore we have the range of T as

$$t_l < T < g^{-1}(\beta_0 + \tilde{\beta}_1 W)$$

Note that since $g()$ is monotone decreasing in this case, g^{-1} has a one-to-one mapping.

When $-\lambda/\beta_2 < 1/\sqrt{2\pi}$, the convergence of expected backlog eq. (EC.11) requires that $\beta_0 + \tilde{\beta}_1 W < g(t^*) = \Phi^{-1}(1-\lambda t^*) - \beta_2 t^*$. At this time $g()$ is inverted U-shaped. Then we will have two intersection points for T. We need

$$g_{left}^{-1}(\beta_0 + \tilde{\beta}_1 W) < T < g_{right}^{-1}(\beta_0 + \tilde{\beta}_1 W)$$

We plot $g(t)$ based on our parameter values in the given range for T. It is monotone decreasing as shown in the figure above.

We see that the first case fits the reality based on the data as shown in Figure EC.2. The intersection point gives us $g^{-1}(\beta_0 + \tilde{\beta}_1 W) \approx 0.264$ which is smaller than $1/\lambda$. Therefore to maintain convergence condition, the average service time should be $t_l < T < g^{-1}(\beta_0 + \tilde{\beta}_1 W)$.

□

After obtaining the stability condition, we further explore the structures of $\rho_i(T)$ and $r(T)$.

Lemma 1. *Suppose the stability condition in Proposition 1 is satisfied, $\tilde{\beta}_1 > 0, \beta_2 < 0$, then both $\rho_i(T)$ and $r(T)$ as defined in Equation* (EC.10) *and eq.* (EC.13) *monotonically increase in T.*

Proof: The first order derivative is as follows:

$$\begin{aligned} r'(T) &= \frac{\lambda\left(1-\Phi\left(\beta_0+\beta_1 W+\beta_2 T\right)\right)-\lambda T\left(-\phi\left(\beta_0+\beta_1 W+\beta_2 T\right)\cdot\beta_2\right)}{\left(1-\Phi\left(\beta_0+\beta_1 W+\beta_2 T\right)\right)^2} \\ &= r(T)/T+\beta_2 r(T) h(\beta_0+\beta_1 W+\beta_2 T) \\ \frac{r'(T)}{r(T)} &= 1/T+\beta_2 h(\beta_0+\beta_1 W+\beta_2 T) \end{aligned}$$

It suffices to focus on $1+\beta_2 T h(\beta_0+\tilde{\beta}_1 W+\beta_2 T)$ where $h(\cdot)$ stands for the hazard function for standard normal distribution: $h(x)=\phi(x)/1-\Phi(x)$.

From eq. (EC.14) we know that when the system is stable, we have $\Phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T)<1-\lambda T$, we then have

$$h(\beta_0+\tilde{\beta}_1 W+\beta_2 T)=\frac{\phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T)}{1-\Phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T)}<\phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T)/\lambda T$$

We only need to see

$$\phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T)<-\lambda/\beta_2.$$

This naturally holds because we have

$$\phi(\beta_0+\tilde{\beta}_1 W+\beta_2 T)\le 1/\sqrt{2\pi}\le -\lambda/\beta_2$$

Therefore we obtain

$$h(\beta_0+\tilde{\beta}_1 W+\beta_2 T)<-\frac{1}{\beta_2 T}. \tag{EC.16}$$

Since $\beta_2<0$, we have

$$\frac{r'(T)}{r(T)}=1/T+\beta_2 h(\beta_0+\beta_1 W+\beta_2 T)\ge 0. \tag{EC.17}$$

Since $r(T)>0$, we obtain that $r'(T)>0$ and thus $r(T)$ is monotonically increasing in T.

Next we will prove $\rho_i'(T)>0$. By eq. (EC.10), we have

$$\frac{d\rho_i}{dT}=\frac{\lambda\cdot\left(1-\Phi(\beta_0+\tilde{\beta}_1 i\cdot T+\beta_2 T)\right)+\lambda T\cdot\phi\left(\beta_0+\tilde{\beta}_1 i\cdot T+\beta_2 T\right)\cdot(\tilde{\beta}_1 i+\beta_2)}{\left(1-\Phi(\beta_0+\tilde{\beta}_1 i\cdot T+\beta_2 T)\right)^2},\ \text{if } iT<W, \tag{EC.18}$$

$$\frac{d\rho_i}{dT}=r'(T),\ \text{if } iT\ge W. \tag{EC.19}$$

In the $iT>W$ case, we have $\frac{d\rho_i(T)}{dT}=r'(T)>0$. So we are left to prove $\frac{d\rho_i(T)}{dT}>0$ when $iT<W$. Since $\tilde{\beta}_1>0$, the numerator of $\frac{d\rho_i}{dT}$ is

$$\begin{aligned} &\lambda\cdot\left(1-\Phi(\beta_0+\tilde{\beta}_1 iT+\beta_2 T)\right)\left(1+T\cdot h\left(\beta_0+\tilde{\beta}_1 iT+\beta_2 T\right)\cdot(\tilde{\beta}_1 i+\beta_2)\right) \\ &>\lambda\cdot\left(1-\Phi(\beta_0+\tilde{\beta}_1 iT+\beta_2 T)\right)\left(1+T\beta_2\cdot h\left(\beta_0+\tilde{\beta}_1 iT+\beta_2 T\right)\right)>0, \end{aligned}$$

where the last inequality follows (EC.16). We have thus proved $\frac{d\rho_i(T)}{dT}>0$ when $iT<W$.

□

The following is the proof of Proposition 2.

Proof. We use a sample-path (coupling) perspective to prove it. The backlog process $B(T)$ can be seen as a birth-death process with the transition-up rate from n to $n+1$:

$$p_n = \frac{\lambda_n}{\lambda_n + \mu_n} = \frac{1}{1 + 1/\rho_n(T)}$$

where we have $\rho_n(T)$ follows what we defined earlier in eq. (EC.10). The transition-down rate is $1 - p_n$.

Now consider an alternative M/M/1 queueing system in which the average service time is increased to $T(1+\delta)$. We then compare the backlog processes in these two systems from an embedded continuous-time Markov chain perspective. Denote the backlog process in the revised system as $\widetilde{B}(t)$.

By Lemma 1, $\rho_i(T)$ increases in T for each given i, and thus $p_n(\cdot)$ also increases in T for this birth-death process. By a coupling argument, $B(\infty) \leq_{s.t.} \widetilde{B}(\infty)$. Therefore $E[\widetilde{B}(T)] \geq E[B(T)]$.

□

We next prove Proposition 3.

Proof: We denote $L(T) := \beta_0 + \tilde{\beta}_1 \min(kT, W) + \beta_2 T$ and take the derivative of T for $E[R(T)]$ to obtain

$$\begin{aligned} E[R(T)]' &= \frac{\partial \sum_{k=0}^{\infty} \Phi(\beta_0 + \tilde{\beta}_1 \min(kT, W) + \beta_2 T)\pi(k,T)}{\partial T} \\ &= \sum_{k=0}^{\infty} \left[\frac{\partial \Phi(L(T))}{\partial T} \pi(k,T) + \Phi(L(T)) \frac{\partial \pi(k,T)}{\partial T} \right] \end{aligned} \tag{EC.20}$$

Denote $R_k(T) = \prod_{i=0}^{k-1} \rho_i(T)$ and $R_0(T) = 1$, we have

$$\pi(k,T) = \frac{R_k(T)}{\sum_{k=0}^{\infty} R_k(T)}, \quad \frac{\pi(k,T)'}{\pi(k,T)} = \frac{R_k'(T)}{R_k(T)} - \frac{\sum_{k=0}^{\infty} R_k'(T)}{\sum_{k=0}^{\infty} R_k(T)}, \quad k \geq 1$$

We then have

$$\pi(k,T)' < \pi(k,T) \cdot \frac{R_k'(T)}{R_k(T)}$$

Focusing on $R_k'(T)/R_k(T)$ we have

$$\frac{R_k'(T)}{R_k(T)} = \sum_{j=0}^{k-1} \frac{\rho_j'(T)}{\rho_j(T)}$$

and

$$\begin{aligned} \frac{\rho_j'(T)}{\rho_j(T)} &= \frac{1}{T} + (\tilde{\beta}_1 j \cdot 1_{\{jT<W\}} + \beta_2) \cdot h(\beta_0 + \tilde{\beta}_1 \min(jT, W) + \beta_2 T) \\ &< \frac{1}{T} - \frac{\tilde{\beta}_1 j \cdot 1_{\{jT<W\}} + \beta_2}{\beta_2 T} = -\frac{\tilde{\beta}_1 j \cdot 1_{\{jT<W\}}}{\beta_2 T} \end{aligned}$$

where the last inequality is due to the property that $h(x)$ can be seen as increasing when x is not too large (< 5) and the condition that

$$\beta_0 + \tilde{\beta}_1 \min(jT, W) + \beta_2 T \leq \beta_0 + \tilde{\beta}_1 W + \beta_2 T < 5.$$

We thus obtain

$$h(\beta_0 + \tilde{\beta}_1 \min(jT, W) + \beta_2 T) \leq h(\beta_0 + \tilde{\beta}_1 W + \beta_2 T)$$

and eq. (EC.16).

Thus we obtain

$$\frac{R_k'(T)}{R_k(T)} \leq \sum_{j=0}^{k-1} \frac{-\tilde{\beta}_1 j \cdot 1_{\{jT<W\}}}{\beta_2 T} = \frac{-\tilde{\beta}_1}{\beta_2 T} \sum_{j=0}^{k-1} j \cdot 1_{\{jT<W\}} \leq \frac{-\tilde{\beta}_1}{\beta_2 T} \cdot \frac{k(k-1)}{2}$$

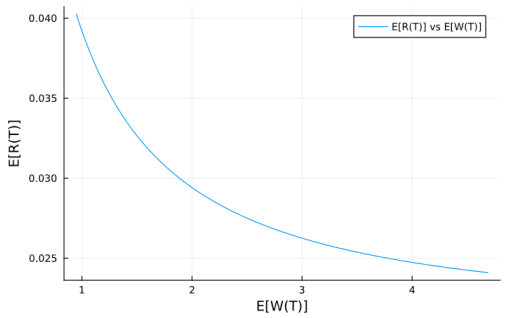

Figure EC.3 Pareto curve between waiting time and reoperation probabilities

Using W/T to upper bound k in ϕ and Φ and denoting $L_W(T) := \beta_0 + \tilde{\beta}_1 W + \beta_2 T$ we obtain

$$\begin{aligned} E[R(T)]' &\leq \sum_{k=0}^{\infty} \left[\phi(L_W(T))(\tilde{\beta}_1 k + \beta_2) + \frac{-\tilde{\beta}_1 \Phi(L_W(T))}{2\beta_2 T} \cdot (k^2 - k) \right] \pi(k, T) \\ &= \beta_2 \phi(L_W(T)) + \tilde{\beta}_1 \left[\phi(L_W(T)) + \frac{\Phi(L_W(T))}{2\beta_2 T} \right] E[B(T)] - \frac{\tilde{\beta}_1 \Phi(L_W(T))}{2\beta_2 T} E[B(T)^2] \\ &= \Phi(L_W(T)) \cdot \left[\mathrm{IMR}(L_W(T))(\beta_2 + \tilde{\beta}_1 E[B(T)]) + \frac{\tilde{\beta}_1}{2\beta_2 T} E[B(T)(1 - B(T))] \right] \end{aligned} \tag{EC.21}$$

Note that the inverse mills ratio (IMR) is defined as $\phi(\cdot)/\Phi(\cdot)$, which is positive and decreasing. When the conditions in Proposition 3 are satisfied, the right-hand-side of (EC.21) is negative (as $\phi(\cdot)$ is positive), proving that the steady-state reoperation rate decreases in T.

EC.4.2. Simulation

For each surgical department, we alter the job allowance for each surgery case by $x\%$ of the current value, where $x \in \{-20, -19, \ldots, 19, 20\}$. For each x, we simulate the surgical waitlist over a 50-year period to achieve the steady state. For some large values of x, the system may be unstable, resulting in an ever-growing waitlist, which will be excluded for further study. For those resulting in a stable queue, we use the last 10 years to compute the steady-state performance metrics, including average wait-to-target ratio and reoperation rate. This process is repeated 50 times for each of the six departments: urology, general, ENT, orthopedics, ophthalmology, plastic. We then simulate the actual surgical durations based on the average estimated effect of slowdown and speed-up while the scheduled shift lengths and EOS times are obtained from original data. The simulated frontiers are displayed in Figure EC.4. The overall framework is:

1. **Parameter setup.** For each department d and each surgeon s, adjust the per-case job allowance by $x\%$ of its baseline value, rounding to the nearest one-minute increment.
2. **Long-run simulation.** Simulate the waitlist over a 50-year horizon to approach steady-state.
3. **Performance measurement.** For each remaining (stable) scenario, compute the average wait-to-target ratio and reoperation rate over the last 10 years of the run.
4. **Replication.** Repeat steps 1–3 for 50 independent replications to obtain Monte Carlo estimates of steady-state metrics for each department and each $x\%$.
5. **Data preparation.**
   - Remove any slowdown/speedup effects from the empirical surgical durations to recover raw case times.
   - In each simulation, compute the actual durations by re-applying the estimated speedup/slowdown factors.
   - Use the original data to generate daily shift schedules for 50 years.

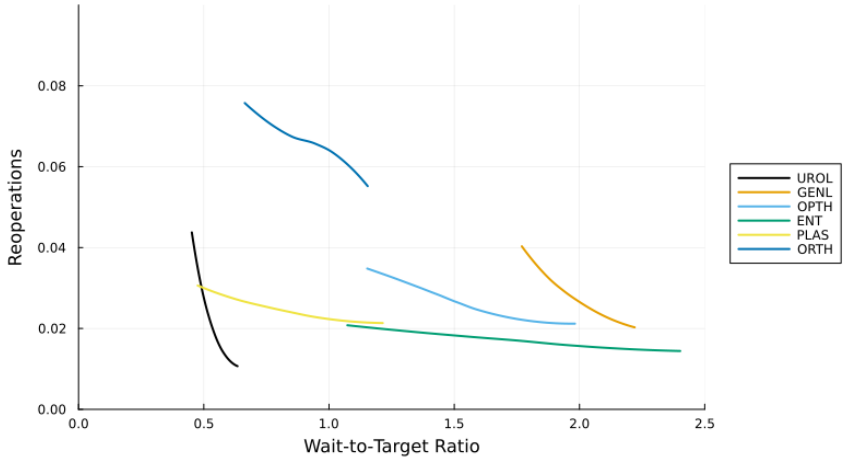

Figure EC.4 All Departments with Unified Scale

Given a focal job allowance level, each simulated operating day proceeds as follows:

Algorithm 1 Daily Waitlist and Reoperation Dynamics

for each surgeon s in the focal department **do**
  Init: current waitlist W, reoperation set R
  for each day $t = 1, \ldots, T$ **do**
    ▷ *Arrivals*: Add new patients to W using the average arrival rate λ_s for the surgeon from data.
    ▷ *Reoperations*: Add any scheduled reoperation cases on day t to W.
    ▷ *Prioritization*: Rank W by wait-to-target ratio, reserving slots for known reoperations.
    ▷ *Scheduling*: Schedule surgeries for surgeon s according to day-t roster, applying the $x\%$ allowance adjustment.
    ▷ *Shift dynamics*: Compute actual service times using the estimated coefficients of adaptive behaviour.
    ▷ *Reoperation draw*: For each completed case, generate a reoperation with probability given by the model; if triggered (probability $> 50\%$), schedule it uniformly 1–30 days ahead and add to R.
  end for
end for