Minimizing Earliness/Tardiness costs on multiple machines with an application to surgery scheduling

Maarten Otten[a*], Aleida Braaksma[a], and Richard J. Boucherie[a]

[a]Center for Healthcare Operations Improvement and Research, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

[*]Corresponding author, e-mail: j.w.m.otten@utwente.nl

Abstract

Early or tardy surgeries are frustrating for both patients and personnel, and cause inefficient use of resources at the operating rooms. The stochastic Earliness/Tardiness (E/T) scheduling problem addresses this by minimizing the total expected deviation of the surgery completion times from the planned completion times. We introduce the concept of E/T-concavity as a property of a probability distribution if the E/T costs are concave as a function of the standard deviation of the completion time, whenever the optimal planned completion times are selected. We use this concept to generate an optimal schedule for the multiple machine variant of the E/T problem. The optimal schedule is not unique and therefore allows us to consider several optimization objectives in addition to the E/T objective. We demonstrate the usefulness of our results in practice by proving E/T-concavity for several probability distributions and by showing that, under the assumption of E/T-concavity, a simple Shortest Variance First (SVF) rule is optimal. We conclude by providing a numerical example of surgery scheduling where we demonstrate the benefits of the SVF rule compared to several commonly used scheduling rules.

Keywords: Surgery scheduling, stochastic scheduling, multiple machine scheduling, Earliness/Tardiness.

1 Introduction

Surgeries that do not start at their scheduled time are a source of frustration for both patients and personnel at the Operating Room (OR) department (Wachtel and Dexter, 2009). Early surgeries cause idle time at the ORs if the next patient is not yet ready for surgery. Tardy surgeries can cause overtime and even cancellation of surgeries. They can also cause tardy starts of subsequent surgeries. Since surgeries usually cause psychological distress, it is undesirable that patients have to wait longer than expected. This is, in particular, inconvenient for patients who are not allowed to eat or drink before surgery. Furthermore, tardy starts result in waiting time for OR personnel and, possibly, other personnel at downstream departments.

The main cause of early and tardy surgeries is that the realized surgery duration differs from the scheduled duration. In part this can be solved by developing methods and models to predict surgery durations more accurately, but due to the stochastic nature of these processes there will always be some variability that has to be taken into account (Eijkemans et al., 2010). This raises the question how the variability of surgery durations is to be taken into account in order to obtain a schedule that, during the day, will deviate as little as possible from the original schedule. In this paper, we aim to improve the robustness of the schedule by considering the expected deviation from the intended schedule in terms of the expected earliness and tardiness of surgeries or appointments. For this we consider the stochastic Earliness/Tardiness (E/T) problem.

Our contribution to literature is threefold: We provide an optimality criterion for surgery scheduling with minimal expected earliness and tardiness, we show how secondary optimization objectives can be incorporated and we show how these results can be implemented in practice. The outline of the remainder of this paper is as follows. First, we give an overview of relevant literature (Section 2). Second, we describe the problem we consider in more detail and introduce the concept of E/T-concavity. We then derive, under the assumption of E/T-concavity, an optimal schedule for the multiple machine E/T problem (Section 3). Third, we consider the E/T problem when additional optimization objectives are taken into account (Section 4). Fourth, we consider the E/T-concavity of several probability distributions and we present an example of how the results of this paper can be applied in practice by implementing a simple Shortest Variance First (SVF) scheduling rule (Section 5). We conclude with a discussion of the results obtained and suggestions for possible future research (Section 6).

2 Literature

Surgery scheduling in general is a widely addressed topic in literature. (Hulshof et al., 2012) and (Cardoen et al., 2010) both provide an extensive overview of literature. However, surgery scheduling which, explicitly, takes variability into account is much less addressed. (Van Riet and Demeulemeester, 2015) and (Ferrand et al., 2014) give an overview of literature where emergency patients are taken into account when the elective patients are scheduled and (Otten et al., 2019) give a model-based overview of literature of surgery scheduling where several types of disturbances are taken into account. In this paper, we aim to improve the robustness of the schedule by considering the expected deviation from the intended schedule in terms of the total expected earliness and tardiness of surgeries. (Denton et al., 2007) consider a similar problem for which they focus on the tardiness of surgeries only. They show that a simple heuristic based on the SVF scheduling rule improves the OR efficiency considerably. (Guda et al., 2016) consider both the earliness and tardiness of surgeries, but only for a single OR. They show that scheduling in increasing order of the surgery duration variance is optimal.

The single machine Earliness/Tardiness (E/T) problem is one of the classical problems in machine scheduling (Pinedo, 2012). It aims to schedule jobs on machines in such a way that the total weighted deviation from their due dates is minimized. The E/T problem also applies to various types of appointments with various resources. Therefore, we use the terms jobs and machines as commonly used in the scheduling literature to introduce our model. The deterministic version of the problem has been widely studied and many of its variants have been proven to be NP-complete (Baker and Scudder, 1990; Hall and Posner, 1991; Hall et al., 1991). More recently, (Soroush and Fredendall, 1994) and (Soroush, 1999) considered the stochastic version. In that version, the processing times of the jobs are assumed to be random. In these papers structural properties were derived and several heuristics to obtain a near-optimal schedule were proposed. These heuristics were later improved in (Portougal and Trietsch, 2006), where also sufficient conditions for an optimal schedule for the variant with distinct Earliness/Tardiness costs parameters were derived. In (Baker, 2014) it is shown, for the single machine E/T problem, that if the processing times are normally distributed it is optimal to schedule jobs in increasing order of variance. This result was extended for generally distributed processing times in (Guda et al., 2016). Several variants of the E/T problem were extended to the multiple machine model, in which jobs are scheduled on multiple parallel identical machines (Cai and Zhou, 1999). In this paper, we consider the stochastic E/T scheduling problem on multiple machines, in which, in addition to a schedule, the due dates also need to be selected. We derive an optimality criterion for this model under the assumption of E/T-concavity and show that an optimal schedule is not unique such that secondary optimization is relevant.

3 The multiple machine Earliness/Tardiness problem

In this section, we describe the variant of the E/T problem, where the jobs are scheduled on multiple machines. We derive several structural properties to obtain conditions for an optimal schedule.

In the Multiple Machine E/T problem, n jobs are scheduled on m identical machines. Let n_i be the number of jobs scheduled on machine i and let $\Pi = [\pi_1 \quad \pi_2 \cdots \pi_m]$ denote the schedule for the m machines for which $\pi_i = (\pi_{i,1} \cdots \pi_{i,n_i})$, $i = 1, 2 \ldots m$, is the sequence in which the jobs are processed at machine i. Further, let $d^{\pi_i} = (d_1^i, \cdots, d_{n_i}^i)$ denote the due dates of the jobs on machine i and P_{ij} the processing time of the job j on machine i. The processing times are assumed to be mutually independent. The completion time of job j on machine i is denoted by

$$C_j^i = \sum_{k=1}^{j} P_{i\pi_{i,k}}.$$

The earliness of job j on machine i is defined as $E_j^i(d_j^i) = (d_j^i - C_j^i)^+$ and its tardiness as $T_j^i(d_j^i) = (C_j^i - d_j^i)^+$. The unit earliness and tardiness cost parameters of job j on machine i are given by α_j^i and β_j^i, respectively. $ET(\Pi, d^\Pi)$ denotes the total expected E/T costs when schedule Π and due dates d^Π, where $d^\Pi = (d^{\pi_1}, \cdots, d^{\pi_m})$, are chosen and the E/T costs of job j on machine i are denoted by

$$F_j^i(d_j^i) = \mathbb{E}\big[\alpha_j^i E_j^i(d_j^i) + \beta_j^i T_j^i(d_j^i)\big], \qquad i = 1, \ldots, m, \quad j = 1, \ldots n.$$

Definition 3.1 *The multiple machine Earliness/Tardiness problem is given by:*

$$(MET) \qquad v^* = \min_{\Pi, d^\Pi} ET(\Pi, d^\Pi) = \min_{\Pi, d^\Pi} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{E}\big[\alpha_j^i E_j^i(d_j^i) + \beta_j^i T_j^i(d_j^i)\big].$$

The problem is termed symmetric if the cost parameters are equal for every job and machine, i.e., $\alpha_j^i = \alpha$ *and* $\beta_j^i = \beta$, *for* $j = 1, \ldots n$

$$(SMET) \qquad v^* = \min_{\Pi, d^\Pi} ET(\Pi, d^\Pi) = \min_{\Pi, d^\Pi} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{E}\big[\alpha E_j^i(d_j^i) + \beta T_j^i(d_j^i)\big].$$

We will omit the superscripts Π and j whenever a result holds for an arbitrary schedule and job, or when Π and j are clear from the context.

Lemma 3.1 *Given a schedule* Π, *the optimal due date of job* j *with completion time* C_j *is* $d_j^* = G^{-1}(\frac{\beta_j}{\alpha_j + \beta_j})$, *with* G *the distribution function of* C_j.

For the proof we refer to (Guda et al., 2016).

Lemma 3.2 *If the optimal due date is chosen, i.e.,* $d_j^* = G^{-1}(\frac{\beta_j}{\alpha_j + \beta_j})$, *then the expected E/T costs can be written as:*

$$F_j^\pi(d_j^*) = \beta_j \int_{d_j^*}^{\infty} x dG(x) - \alpha_j \int_{-\infty}^{d_j^*} x dG(x). \tag{1}$$

Proof. We split the expression for the Earliness and the Tardiness and consider the terms separately:

$$F_j^\pi(d_j^*) = \mathbb{E}\Big[\alpha_j\big(d_j^* - C_j\big)^+ + \beta_j\big(C_j - d_j^*\big)^+\Big], \tag{2}$$

$$\begin{aligned}
\mathbb{E}\Big[\big(d_j^* - C_j\big)^+\Big] &= d_j^* G(d^*) - \int_{-\infty}^{d_j^*} x dG(x) \\
&= d_j^* \frac{\beta_j}{\alpha_j + \beta_j} - \int_{-\infty}^{d_j^*} x dG(x), \\
\mathbb{E}\Big[\big(C_j^i - d_j^*\big)^+\Big] &= \int_{d_j^*}^{\infty} x dG(x) - d_j^* \frac{\alpha_j}{\alpha_j + \beta_j}.
\end{aligned}$$

Substituting this into equation (2) gives the desired result. ■

Note that the optimal due date itself depends on α_j and β_j, and also on the distribution of the completion time. So, $F_j^\pi(\cdot)$ is a function of α_j, β_j and the parameters of the distribution of the completion time.

Definition 3.2 (E/T-concavity) *A probability distribution is E/T-concave if the E/T cost function, $F_j^\pi(d_j^*)$, is concave in the standard deviation of the completion time, σ, where d_j^* is the optimal due date, $d_j^* = G_j^{-1}(\beta_j/(\alpha_j + \beta_j))$.*

E/T-concavity allows us to determine an optimal solution to the E/T problem and holds for several probability distributions, as we will show in Section 5. We now use this concept of E/T-concavity to derive necessary and sufficient conditions for an optimal schedule. For the single machine variant of this problem (Guda et al., 2016) proved that it is optimal to schedule the jobs in increasing order of variance. Their result holds for general distributions and does not depend on the concept of E/T-concavity. In the next theorem we show that an extension of the scheduling in increasing order of variance rule is also optimal for the symmetric multiple machine variant.

Theorem 3.1 (Optimal SMET schedule) *For the SMET problem with E/T-concave completion times, the optimal schedule is such that there are no jobs p_a and q_b with:*

$$Var(P_{a,p}) > Var(P_{b,q}) \qquad \text{and} \qquad n_a - \pi_{a,p} > n_b - \pi_{b,q}, \tag{3}$$

i.e., the number of successors of job p_a on machine a is higher than that of q_b on machine b and the variance of the processing time of job p_a is higher than the variance of the processing time of job q_b. Moreover, every schedule that satisfies this condition is optimal and has the value v^*.

Proof. The proof consists of two parts: first, we prove that every schedule that does not satisfy the condition is non-optimal and second, we prove that every schedule that does satisfy condition (3) is equivalent, i.e., has a value of $v = v^*$.

Suppose that there is an optimal schedule Π such that there are jobs that do not follow condition (3). Let job k on machine a and job $n_b - l$ on machine b be such that $n_a - k > l$ and Var$(P_{a,k})$ >Var(P_{b,n_b-l}). Let $\tilde{\Pi}$ be the schedule in which these two jobs are swapped. The jobs in this new schedule can be divided into three sets, namely (**I**), (**II**) and (**III**), see Figure 1.

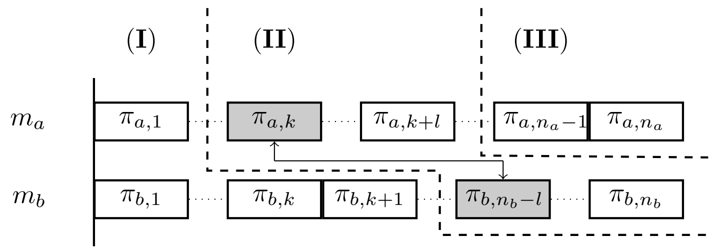

Figure 1: Schedule Π of machines m_a and m_b.

For the jobs in set (**I**), nothing changes and therefore these jobs have the same expected Earliness/Tardiness costs.

For the jobs in set (**II**), define $x =\text{Var}(P_{a,k})-\text{Var}(P_{b,n_b-l})$ and again let $C_{i,j}$ denote the completion time of job j on machine i and $\tilde{C}_{i,j}$ the completion time in the swapped schedule. For $z = 0,\ldots,l$ we have:

$$\begin{aligned} \text{Var}(\tilde{C}_{a,k+z}) &= \text{Var}(C_{a,k+z}) - x \\ \text{Var}(\tilde{C}_{b,n_b-l+z}) &= \text{Var}(C_{b,n_b-l+z}) + x \end{aligned} \tag{4}$$

The expected E/T costs (F^i_j) are concave as a function of the standard deviation σ by assumption of E/T-concavity. Since the function $h(x) = \sqrt{x}$ is also concave, the expected E/T costs are also concave as a function of the variance σ^2. We therefore obtain:

$$\begin{aligned} &F^a_{k+z}(\text{Var}(\tilde{C}_{a,k+z})) + F^b_{n_b-l+z}(\text{Var}(\tilde{C}_{b,n_b-l+z})) \\ &= F^b_{n_b-l+z}(\text{Var}(C_{b,n_b-l+z} + x)) + F^a_{k+z}(\text{Var}(C_{a,k+z} - x)) \\ &\geq F^b_{n-l+z}(\text{Var}(C_{b,n_b-l+z})) + F^a_{k+z}(\text{Var}(C_{a,k+z})). \end{aligned} \tag{5}$$

Also, from (4) and concavity of F we have:

$$\begin{aligned} &F^a_{k+z}(\text{Var}(C_{a,k+z})) + F^b_{n_b-l+z}(\text{Var}C_{b,n_b-l+z})) \\ &= F^b_{n_b-l+z}(\text{Var}(\tilde{C}_{b,n_b-l+z} - x)) + F^a_{k+z}(\text{Var}(\tilde{C}_{a,k+z} + x)) \\ &\geq F^b_{n_b-l+z}(\text{Var}(\tilde{C}_{b,n_b-l+z})) + F^a_{k+z}(\text{Var}(\tilde{C}_{a,k+z})). \end{aligned} \tag{6}$$

Combining (5) and (6) gives:

$$F^b_{n_b-l+z}(\text{Var}(C_{b,n_b-l+z})) + F^a_{k+z}(\text{Var}(C_{a,k+z})) = F^b_{n_b-l+z}(\text{Var}(\tilde{C}_{b,n_b-l+z})) + F^a_{k+z}(\text{Var}(\tilde{C}_{a,k+z})), \quad z = 0,\ldots,l.$$

Therefore, the total expected E/T costs for the jobs in set (**II**) remain unchanged.

The set of jobs (**III**) is nonempty because $n_a - k > l$ and the variance of these jobs is reduced by $x =\text{Var}(P_{a,k})-\text{Var}(P_{b,n_b-l}) > 0$. Therefore, the total expected E/T costs must be lower since F is an increasing function in the variance.

In conclusion, we have that the total expected E/T costs of the swapped schedule are less than those of the original schedule; this contradicts our assumption that the original schedule is optimal.

For the second part of the proof, note that every schedule that satisfies condition (3) can be obtained from any other schedule that satisfies condition (3) by a finite number of swaps. Furthermore, these swaps are all of the type of swapping the $(n_a - i)^{th}$ job on machine a with the $(n_b - i)^{th}$ job on machine b. Therefore, using the same line of reasoning as in the first part of the proof, we see that the total expected Earliness/Tardiness costs are equal since in this case the set (**III**) is empty.

In conclusion, we showed that every schedule that does not satisfy the increasing variance condition (3) is non-optimal and that every schedule that does, has the same value of v^*, hence every increasing variance schedule is optimal. ■

Theorem 3.1 provides a criterion for a schedule to be optimal. From this, an optimal schedule can be readily obtained. In case that the number of jobs on machine a exceeds the number of jobs on machine b by more than one, Theorem 3.1 implies that it is better to move the job with the smallest variance on machine a to machine b. This can be seen as interchanging the job on machine a with a zero variance dummy job on machine b. To construct an optimal schedule, let $h = n \mod m$, there are h machines with $\lceil \frac{n}{m} \rceil$ jobs and $m - h$ machines with $\lfloor \frac{n}{m} \rfloor$ jobs. In an optimal schedule the h jobs with the smallest variance are scheduled on the h machines with the most jobs and the other jobs are divided in sets of m jobs based on their variance. For this, assume that the jobs $i = 1, \ldots, n$ are ordered in increasing order of variance, i.e., $\sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_n$. Let $k = \lceil \frac{n}{m} \rceil$ and define:

$$\begin{aligned} &\mathcal{I}_1 = \big\{(1, \ldots, h\big\}, \\ &\mathcal{I}_j = \big\{(j-2) * m + h + 1, \ldots, (j-2) * m + h + m\big\}, \qquad\qquad j = 2, \ldots, k. \end{aligned} \tag{7}$$

Sets $\mathcal{I}_2, \ldots, \mathcal{I}_k$ each contain m jobs and the variance of a job in set $\mathcal{I}_j$ is smaller than or equal to the variance of a job in set $\mathcal{I}_l$ whenever $j < l$. To construct an optimal schedule, we assign exactly one job from set $\mathcal{I}_2, \ldots, \mathcal{I}_k$ to each machine, such that a job from set j has ranking number $j - 1$, and finally, assign the jobs from set 1 to h machines as first job, thereby increasing the ranking numbers of the other jobs on the machine by one.

4 Secondary optimization

The results of Section 3 provide a straightforward way to construct a schedule for which the expected earliness and tardiness of the surgeries is minimal. However, in practice minimizing the expected earliness and tardiness is not the only goal of a hospital. In addition to sticking to the original schedule there are various other objectives that a hospital aims for when making a surgery schedule. (Cardoen et al., 2010) list over 20 different objectives, like OR utilization, overtime and makespan, that are used in the surgery scheduling literature. Fortunately, E/T-optimal schedules are not unique. Interchanging surgeries with the same ranking number at different ORs does not affect the total expected E/T costs. With n surgeries in m ORs, there are $\big((m!)^{k-1}\big) \cdot ((n \bmod m)!)$ E/T-optimal schedules, because for each of the sets $\mathcal{I}_2, \ldots, \mathcal{I}_k$, $k = \lceil \frac{n}{m} \rceil$ there are $m!$ possibilities to allocate the surgeries to the ORs and for the set $\mathcal{I}_1$ there are $(n \bmod m)!$ possibilities. This leaves room for secondary optimization. For example, if the left schedule in Figure 2 is optimal with respect to the E/T objective, swapping surgery 3 and 4 does not affect the total expected E/T costs but results in a schedule with minimal makespan. In this section we will consider secondary optimization objectives that are relevant for surgery scheduling.

4.1 Makespan optimization

The makespan of a schedule is the maximal expected completion time of the surgeries. To determine a schedule for which the makespan is minimal we can formulate the problem as an Integer Linear Program (ILP). Let x_{ij} denote the binary variable that is 1 when surgery j is scheduled in OR i and 0 otherwise. Let m denote the number of ORs and n the number of surgeries. The expected duration of surgery i is $\mathbb{E}(P_i)$. The ILP that minimizes the makespan

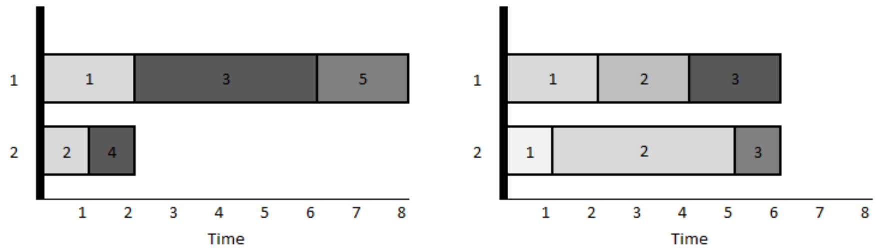

Figure 2: Example of secondary makespan optimization.

is:

$$
\begin{aligned}
(MS)\ \min \quad & y \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_{ij}\mathbb{E}(P_i) \leq y, && i = 1, \ldots, m, \\
& \sum_{i=1}^{m} x_{ij} = 1, && j = 1, \ldots, n, \\
& x_{ij} \in \{0, 1\} && i = 1, \ldots m, \quad j = 1, \ldots, n. \qquad (8)
\end{aligned}
$$

Makespan minimization is an NP-hard problem (Lenstra et al., 1977; Garey and Johnson, 1978), hence the above ILP cannot be solved efficiently for large instances. Then heuristics are needed, like scheduling the surgeries in decreasing order of expected durations (Pinedo, 2012).

Minimizing the makespan in the set of E/T-optimal schedules has less freedom in assigning surgeries to ORs than the general case. In the previous section we described a method to obtain an optimal E/T schedule, see (7). Using this notation, the ILP for this adjusted makespan problem is:

$$
\begin{aligned}
(MS/ET)\ \min \quad & y \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_{ij}\mathbb{E}(P_i) \leq y, && i = 1, \ldots, m, \\
& \sum_{i=1}^{m} x_{ij} = 1, && j = 1, \ldots, n, \\
& \sum_{j \in \mathcal{I}_l} x_{ij} = 1, && i = 1 \ldots m, \quad l = 1, \ldots, k, \qquad (9) \\
& x_{ij} \in \{0, 1\}, && i = 1 \ldots m, \quad j = 1, \ldots, n. \qquad (10)
\end{aligned}
$$

To ensure that a feasible solution exists we add dummy jobs with zero mean and variance to the set $\mathcal{I}_1$, such that $|\mathcal{I}_1| = m$. This ensures that constraints (9) are satisfied. Unfortunately, optimizing over this set of possible schedules is still NP-hard. We can show this by reducing the problem to the bin-packing problem, which is known to be NP-hard (Pinedo, 2012). There are several good heuristics to solve this problem efficiently. A simple heuristic is a greedy algorithm that assigns surgery $i \in \mathcal{I}_j$ to the OR with the least load which does not already have a surgery from the set $\mathcal{I}_j$ assigned to it.

As an illustration of secondary optimization we consider $n = 25$ surgeries and $m = 5$ ORs, where we have randomly drawn the mean and standard deviation from uniform distributions. The surgeries, denoted as [mean, standard deviation], are: [3, 0.25], [2, 0.47], [4, 0.2], [1, 0.75], [21, 0.93], [21, 0.75], [3, 0.29], [1, 0.47], [2, 0.57], [4, 0.08], [3, 0.38], [4, 0.06], [20, 0.78], [1, 0.34],

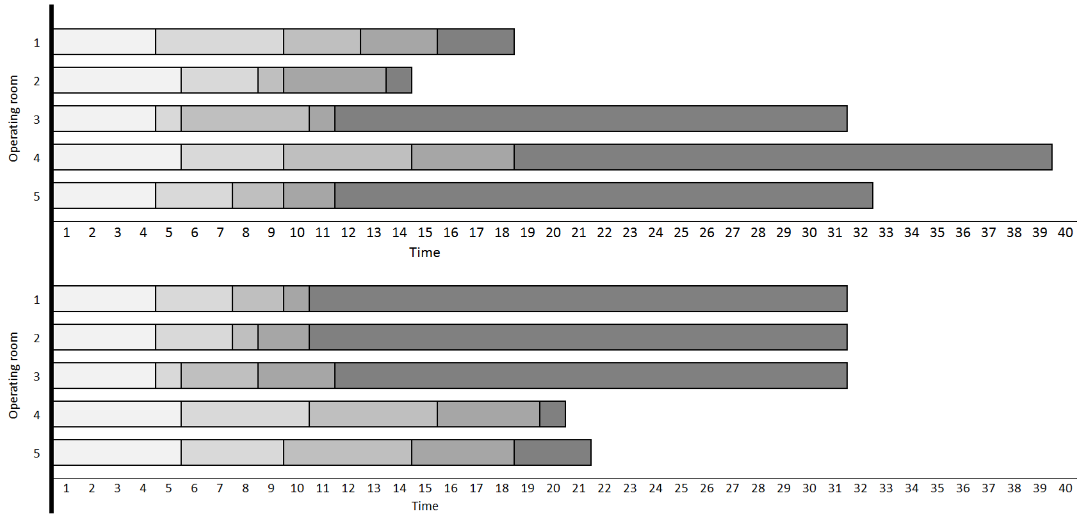

Figure 3: Top: minimal expected E/T costs. Bottom: Minimal expected E/T costs with secondary makespan optimization. The length of a bar corresponds with the expected surgery duration and the color with the variance (darker is higher variance).

[3, 0.61], [5, 0.02], [4, 0.57], [1, 0.83], [5, 0.53], [5, 0.35], [4, 0.58], [5, 0.13], [4, 0.35], [3, 0.91] and [5, 0.55].

In Figure 3 the surgeries are first scheduled such that the E/T costs are minimized, and second such that the makespan is also minimized. We can see that while the total expected E/T costs remain the same, the expected makespan reduces from 39 to 31.

Makespan optimization also leaves room for secondary optimization. The makespan is determined by allocation of the surgeries to the ORs and is not affected by the schedules on the individual ORs. So, just as we can apply makespan optimization as a secondary optimization objective, we can first optimize with respect to the the makespan objective and secondary with respect to the E/T objective. Figure 4 shows an example, with the same surgeries as before, of primary makespan optimization and secondary E/T optimization. The diagram shows that the makespan remains minimal whereas the total E/T costs decrease from 26.5 to 24.2. The minimal makespan can be determined by the ILP described in (8). This assigns each surgery to an OR after which it is optimal to sequence them in increasing order of variance (Guda et al., 2016).

Table 1 shows the optimal values for the expected E/T costs (v^*) and makespan (M^*). As we can see, the order of optimization matters. Primary E/T optimization and secondary makespan optimization yields a reduction of more than 20% of the makespan, whereas the reverse order of optimization results in a reduction of less than 10% of the E/T costs. So depending on the scheduler's judgment which objective is most important, the order of optimization should be

Primary objective	E/T	E/T	Makespan	Makespan
Secondary objective	-	Makespan	-	E/T
v^*	21	21	26.5	24.2
M^*	39	31	27	27

Table 1: Comparison of the order of optimization. The optimal E/T costs are denoted by v^* and the optimal makespan by M^*.

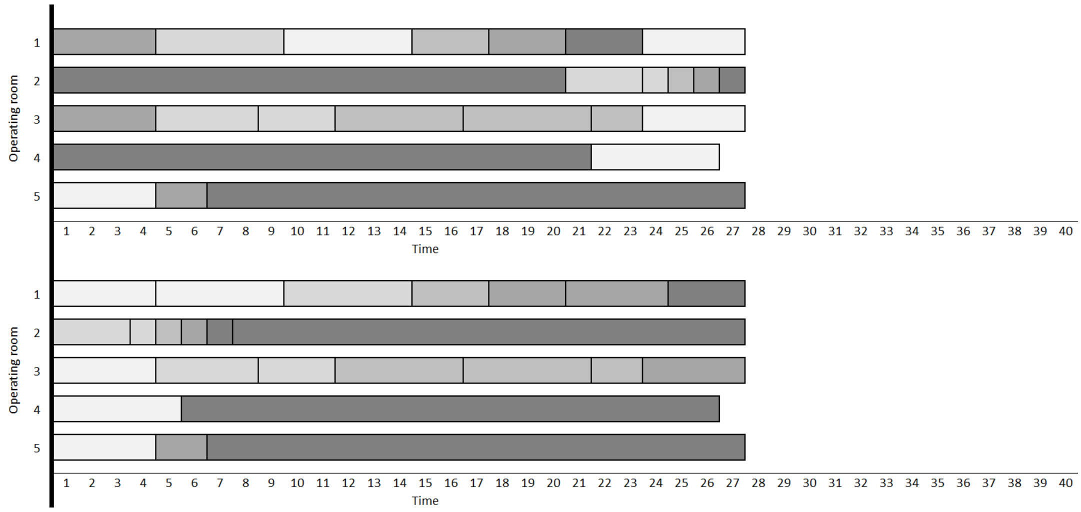

Figure 4: Top: minimal expected makespan. Bottom: Minimal expected makespan with secondary E/T optimization. The length of a bar corresponds to the expected surgery duration and the color with the variance (darker is higher variance).

determined.

4.2 Overtime optimization

Another useful objective for secondary optimization is overtime minimization. With this objective we aim to minimize the probability that an OR and its personnel have to work in overtime, i.e., that an OR is still used past the scheduled completion time of the last surgery. In order to minimize this probability, the variance of the last surgery's completion time has to be minimized. Because the surgery durations are all mutually independent, the variance of the last surgery's completion time is the sum of the variances of all the surgery durations in an OR. So, in fact, overtime minimization is equivalent to makespan minimization, where the sum of the mean surgery durations is minimized instead of the sum of the variances. Consequently, we can use, mutatis mutandis, the ILP of the previous section to solve this problem.

5 Results

In this section we present several results for the E/T problem that demonstrate its usefulness in practice. First, we show that the normal and exponential distributions are E/T-concave. Second, we discuss E/T-concavity for several other distributions. Finally, we demonstrate the applicability of these results for a practical surgery scheduling example.

5.1 E/T-concavity

The optimality of a schedule generated using Theorem 3.1 requires the E/T-concavity of the distribution of the completion times. In this section we consider the E/T-concavity of several probability distributions. First we show that the normal distribution is E/T-concave, which is a very useful result since the sum of independent normally distributed random variables is again normally distributed. So if the processing times are normally distributed, then the completion times will be too.

Lemma 5.1 *The normal distribution is E/T-concave.*

Proof. We have to show that the E/T cost function

$$F(d^*) = \alpha\mathbb{E}\left[(d^* - C)^+\right] + \beta\mathbb{E}\left[(C - d^*)^+)\right],$$

where $d^* = G^{-1}(\beta/(\alpha+\beta))$, is concave in the standard deviation, σ, of the completion time C. The completion time C is normally distributed with mean μ and variance σ^2, so the probability density function is:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

By definition, the due date is fixed at the optimal value (see Lemma 3.1), so we now consider the E/T costs as a function of σ. Recalling Lemma 3.2, evaluating the integrals separately gives:

$$\begin{aligned}\int_{d^*}^{\infty} xdG(x) &= \int_{d^*}^{\infty} x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\ &= \sigma^2\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(d^*-\mu)^2}{2\sigma^2}} + \mu\frac{\alpha}{\alpha+\beta} \\ &= \sigma^2 g(d^*) + \mu\frac{\alpha}{\alpha+\beta}\end{aligned}$$

and similarly:

$$\int_{-\infty}^{d^*} xdG(x) = -\sigma^2 g(d^*) + \mu\frac{\beta}{\alpha+\beta}.$$

Combining the two expressions gives:

$$F(\sigma) = (\alpha+\beta)\sigma^2 g(d^*).$$

Standardization then gives $d^* = G^{-1}(\beta/(\alpha+\beta)) = \mu + \sigma\Phi^{-1}(\beta/(\alpha+\beta))$, where Φ is the distribution function of a standard normal distribution. Therefore $\Phi^{-1}(\beta/(\alpha+\beta))$ is independent of μ and σ and is constant for fixed parameters α and β. Define $q = \Phi^{-1}(\beta/(\alpha+\beta))$, rewriting the derived expression gives:

$$\begin{aligned}F(\sigma) &= (\alpha+\beta)\sigma^2 g(d^*) \\ &= (\alpha+\beta)\sigma^2 g(\mu+\sigma q) \\ &= \sigma\frac{(\alpha+\beta)}{\sqrt{2\pi}}e^{-\frac{1}{2}q^2}.\end{aligned}$$

This is linear in σ, so we can conclude that the normal distribution is E/T-concave. ■

Proving that a probability distribution is E/T-concave is rather involved for general distributions. The hard part, generally, is that there is no tractable or even closed form for the optimal due date d^*, which is determined by the inverse cumulative distribution function (quantile function). However, for the exponential distribution the quantile function is simple.

Lemma 5.2 *The exponential distribution is E/T-concave.*

Proof. The proof follows the lines of that of Lemma 5.1. We have:

$$F(\sigma) = \beta\Big(d^* + \sigma\Big)e^{-\frac{1}{\sigma}d^*} - \alpha\Big[\sigma - \Big(d^* + \sigma\Big)e^{-\frac{1}{\sigma}d^*}\Big].$$

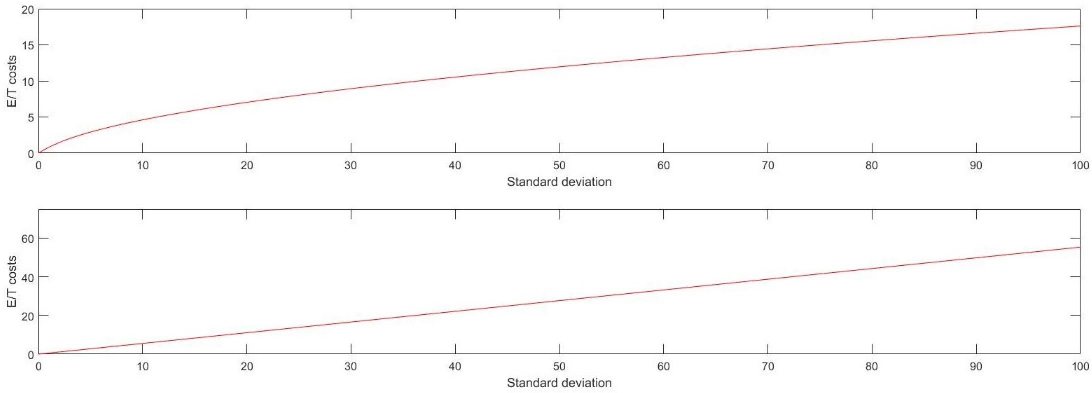

Figure 5: Simulated expected E/T costs as a function of the standard deviation for $C \sim \text{Lognormal}(\nu, \tau)$ with ν fixed (top) and τ fixed (bottom).

Substituting $d^* = G^{-1}(\frac{\beta}{\alpha+\beta})$ with G^{-1} the quantile function of the exponential distribution gives:

$$F(\sigma) = \left(\frac{\alpha(\alpha+\beta)}{\alpha+\beta} \left[1 - \ln(\frac{\alpha}{\alpha+\beta}) \right] - 1 \right) \sigma,$$

which is linear in the standard deviation, so the exponential distribution is E/T-concave. ■

Recall from Definition 3.2 that a probability distribution is E/T-concave if the E/T cost function is concave in the standard deviation of the completion time, which is a sum of processing times. Since there exist no probability distributions for the processing times such that their sums is exponentially distributed, the E/T-concavity of the exponential distribution has no immediate practical relevance. It does, however, show that E/T-concavity is not limited to the normal distribution only, and gives reason to investigate this property for more probability distributions. For the lognormal distribution we simulated the expected E/T costs for several values of the standard deviation for the case that $\alpha = \beta = 1$. The results of this simulation are shown in Figure 5. The simulation results suggest that this probability distribution is also E/T-concave. Note that for lognormally distributed processing time the completion times in general are not lognormally distributed, because the sum of lognormal random variables is not lognormal. However it can be approximated reasonably well by a lognormal distribution (Mehta et al., 2007). This suggests that the optimal schedule for E/T-concave distributions will also be near-optimal for lognormal processing times. We will show this in the next section.

5.2 Application to surgery scheduling

In order to show how the results from this paper can be applied in practice and to compare the performance of several schedules from an E/T perspective, we consider an instance of an OR scheduling problem schedule in this section. The instance is obtained from a benchmark set for surgery scheduling (Leeftink and Hans, 2018). This benchmark set is based on real-life data obtained from several hospitals in The Netherlands. The instance is generated as a mix of surgeries from all specialties and for 10 ORs that are open for 8 hours a day. We specify the surgeries by their expected duration, μ, and standard deviation, σ. The complete instance of 53 surgeries is given in Table 2.

We consider both normally and lognormally distributed surgery durations. Normally distributed surgery durations are commonly used in literature and they imply that the completion times are also normally distributed (Marcon et al., 2003; Van Houdenhoven et al., 2013). Since

Surgery	μ	σ	Surgery	μ	σ	Surgery	μ	σ
1	204	196.1	19	85	19.6	37	85	19.6
2	32	9.6	20	143	98.6	38	86	71.1
3	86	71.1	21	73	15.7	39	61	13.1
4	49	32.3	22	86	71.1	40	85	19.6
5	48	11.7	23	37	27.1	41	204	196.1
6	204	196.1	24	85	50.8	42	86	71.1
7	85	50.8	25	298	52.3	43	37	10.7
8	86	71.1	26	26	9.2	44	73	15.7
9	86	71.1	27	204	196.1	45	156	58.6
10	143	98.6	28	32	9.6	46	20	6.1
11	25	11.7	29	85	50.8	47	73	15.7
12	68	23.6	30	86	71.1	48	17	7.3
13	156	195.1	31	61	20.4	49	61	13.1
14	85	19.6	32	21	7.7	50	49	32.3
15	268	98.8	33	143	98.6	51	81	59.4
16	85	50.8	34	32	9.6	52	144	47.9
17	47	14.7	35	30	11.4	53	85	50.8
18	85	19.6	36	30	7.9	**Av.**	**90.6**	-

Table 2: Surgery scheduling instance. μ (mean) and σ (standard deviation) are in minutes.

we have shown that the normal distribution is E/T-concave, we know that scheduling in increasing order of variance is optimal in this case. Furthermore, by using the expression from Lemma 5.1, the E/T costs can be calculated explicitly. We compare this with lognormally distributed surgery durations because literature suggests that the lognormal distribution is better suited for surgery durations than the normal distribution (May et al., 2000; Eijkemans et al., 2010). Also, (Leeftink and Hans, 2018) assume a lognormal distribution for the surgery durations.

The top row of Table 3 shows the total and average earliness and tardiness (in minutes) for the instance, under the assumption of normally distributed surgery durations. We compare the total expected E/T costs of the optimal Smallest Variance First (SVF) schedule with a random schedule, the Shortest Surgeries First (SSF) schedule and the Longest Surgeries First (LSF) schedule. We choose the SSF and LSF rules because they are commonly used in practice (Pinedo, 2012), whereas the random schedule serves as a reference point. We notice that, for normally distributed surgery durations, the SVF rule indeed performs best. Note that the SSF rule performs almost as well as the SVF rule and that LSF performs worse than random. A plausible reason for the good performance of the SSF rule is that short surgeries tend to have a smaller variance than long surgeries so that the SSF schedule is very similar to the SVF schedule. The second row of Table 3 shows the 99% confidence intervals of total and average E/T costs for the four scheduling rules when the surgery durations are lognormally distributed. In order to obtain the confidence intervals we simulated realizations of the schedule. We draw 100,000 realizations from 53 lognormal random variables with parameters as in Table 2. We use a Common Random Numbers approach (i.e., we use the same random number stream for simulating each of the scheduling rules) to ensure the validity of the comparison of the different scheduling rules. For the lognormal distribution we see that the performance of the scheduling

	SVF	Random	SSF	LSF
Normal				
($\alpha = 1, \beta = 1$)				
Total E/T costs	2563	5545	2690	6666
Average E/T costs	48.4	104.6	50.8	125.8
Lognormal				
($\alpha = 1, \beta = 1$)				
Total E/T costs	2286 $\pm$ 13.7	4616 $\pm$ 46.0	2364 $\pm$ 14.6	5568 $\pm$ 53.1
Average E/T costs	43.1 $\pm$ 0.3	87.1 $\pm$ 0.9	44.6 $\pm$ 0.3	105.1 $\pm$ 1.0
($\alpha = 4/3, \beta = 2/3$)				
Total E/T costs	1913 $\pm$ 10.3	3758 $\pm$ 33.2	1964 $\pm$ 10.9	4547 $\pm$ 38.6
Average E/T costs	36.1 $\pm$ 0.2	71.0 $\pm$ 0.6	37.1 $\pm$ 0.2	85.8 $\pm$ 0.7
($\alpha = 2/3, \beta = 4/3$)				
Total E/T costs	2292 $\pm$ 15.9	4758 $\pm$ 55.0	2386 $\pm$ 16.9	5721 $\pm$ 63.4
Average E/T costs	43.2 $\pm$ 0.3	89.8 $\pm$ 1.0	45.0 $\pm$ 0.3	108.0 $\pm$ 1.2

Table 3: The total and average earliness and tardiness (in minutes) for normally distributed surgery durations and 99% confidence intervals for the total and average expected earliness and tardiness (in minutes) for lognormally distributed surgery durations for several scheduling rules (SVF = Smallest Variance First, SSF/LSF = Shortest/Longest Surgery First).

rules is very similar to the performance under the normal distribution. The results of this simulation together with the E/T-concavity simulation suggest that also for lognormal surgery durations the SVF rule is optimal. Figure 6 shows the earliness and tardiness for each of the surgeries for the different scheduling rules, if we assume a lognormal distribution. We see that for the SVF and SSF rule more than half of the surgeries do not deviate more than 30 minutes from their scheduled start times whereas with the LSF and Random rule more than half of the surgeries deviate more that 90 minutes from their scheduled start times.

The average E/T costs per surgery are more than half of the average surgery duration (90.6 minutes, Table 2). This very high average deviation from the planned schedule is due to the fact that our instance contains many highly variable surgeries and the fact that idle time between surgeries is not allowed. Due to the lack of idle time there is no buffer that can absorb large deviations from the schedule. If we assume that there is idle time halfway the day, e.g. a lunch break, to perform as a buffer, our simulation shows that the expected E/T costs are reduced by almost 10%.

We see that the total expected E/T costs are over 50% lower for the SVF schedule compared to the random schedule and 60% lower compared to the LSF schedule. If the E/T costs are not taken into consideration, the planned schedule may be very efficient but the effectuated schedule is more likely to deviate from it, with waiting patients and hospital personnel as a result.

For the results above we set the E/T costs parameters $\alpha = \beta = 1$, i.e., earliness and tardiness of surgeries are regarded equally undesirable. The bottom two rows of Table 3 show the 99% confidence intervals for two cases where $\alpha \neq \beta$. We see that if earliness is regarded twice as undesirable as tardiness the E/T costs decrease whereas if tardiness is regarded twice as undesirable as earliness the E/T costs increase, compared to the case where $\alpha = \beta$. This is due to the fact that the lognormal distribution is right-skewed, making the earliness of a job

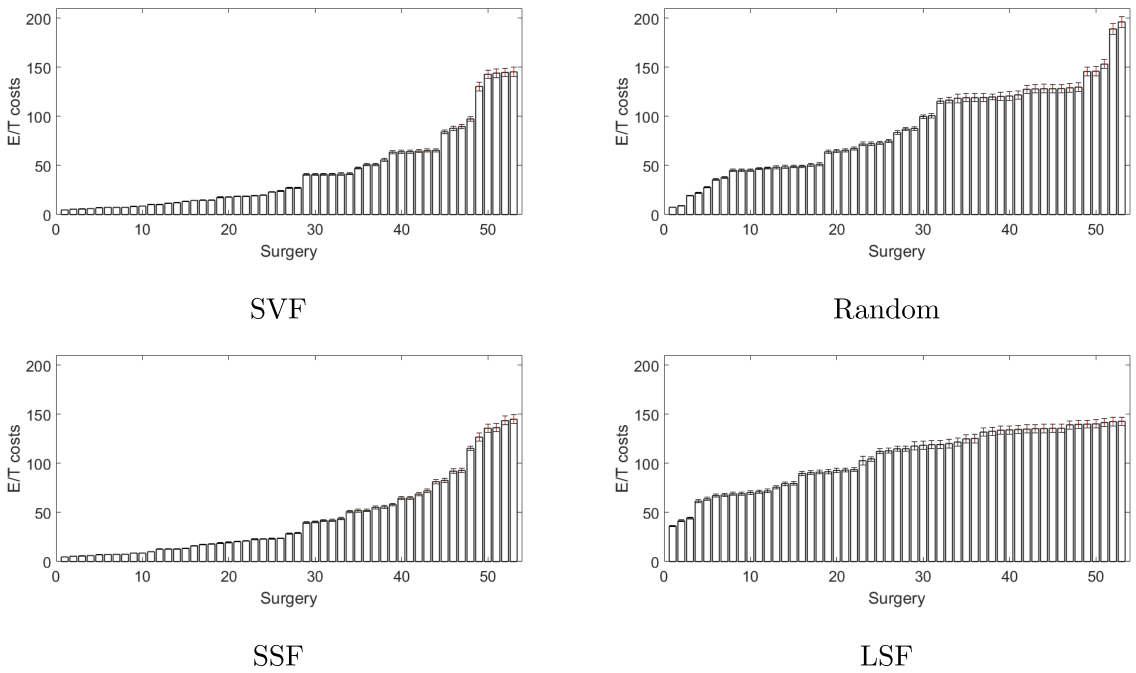

Figure 6: Earliness and tardiness (in minutes) per surgery (99% confidence intervals) for lognormal distributed durations, with $\alpha = \beta = 1$, for several scheduling rules (SVF = Smallest Variance First, SSF/LSF = Shortest/Longest Surgery First).

less variable than its tardiness. Figure 7 shows the total E/T costs of the SVF schedule for the normal distribution and lognormal distribution for various values of α and β. We see that for the normal distribution, which is not skewed, the graph is symmetric with the maximum at $\alpha = \beta = 1$. For the lognormal distribution, which is right-skewed, it is slightly asymmetric with the maximum at $\alpha < \beta$.

6 Discussion

In this paper we considered the problem of early and tardy surgeries. We derived an optimality criterion for a schedule to have minimal total expected earliness and tardiness, for the case that the distribution of the completion times is E/T-concave. We showed that the normal and the exponential distribution are E/T-concave. Computer simulation suggests that this property also holds for the lognormal distribution. Therefore, it would be interesting to see if this conjecture holds and to study E/T-concavity in greater depth by investigating under which conditions this property holds for a probability distribution. With this the optimality of the SVF schedule could be extended to more general distributions of the completion times.

Optimizing with respect to the E/T objective leaves room for additional optimization, because interchanging surgeries with the same rank number does not affect the total expected E/T costs. In this paper we investigated two secondary optimization objectives, namely, makespan and overtime minimization. Other objectives applicable to surgery scheduling might be explored, as well as the optimal order of optimization.

The SVF-rule, which we showed to be optimal under the assumption of E/T-concavity, is easy to implement for schedulers. Also if other objectives are deemed more important, the SVF rule can be applied to the parts of the schedule where the order of surgeries is not fixed, e.g., blocks of surgeries from one specialty, in order to make the schedule more robust. For a practical surgery scheduling example, we demonstrated that the SVF rule performs substantially better

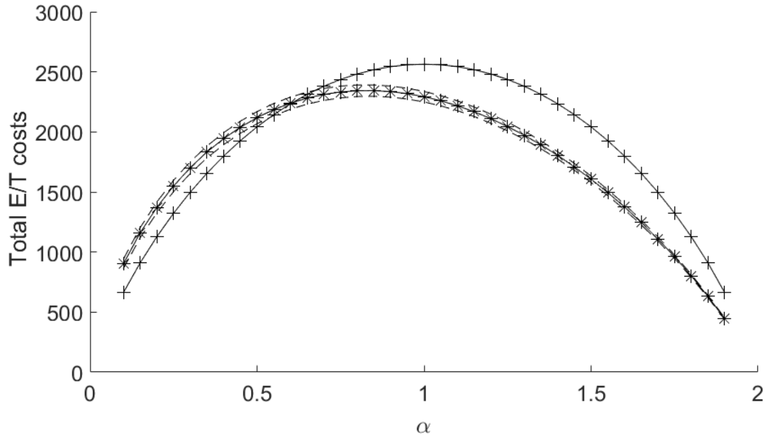

Figure 7: The total E/T costs of the SVF schedule with normal (+) and lognormal (*, with 99% confidence interval) distributed surgery durations for $0 < \alpha < 2$ and $\beta = 2 - \alpha$.

than other commonly used scheduling rules, also for lognormally distributed surgery durations. This is an important insight because it is common practice in hospitals to schedule long surgeries at the beginning of the day (Hans et al., 2008). The rationale behind this is that the schedulers prefer short, low-variable surgeries at the end of the day such that, by late cancellations, they can prevent overtime. However, since surgeries with long expected durations also tend to have higher variances, this results in a schedule for which the actual starting times of the surgeries will deviate more from the planned starting times.

SVF, as a rule of thumb, improves the robustness of a surgery schedule. Scheduling low-variable surgeries at the start and high-variable surgeries at the end of the day, increases the reliability of the planned start times and reduces frustration of unnecessary waiting time for both patients and personnel.

References

Baker, K. R. (2014). Minimizing earliness and tardiness costs in stochastic scheduling. *European Journal of Operational Research*, 236(2):445–452.

Baker, K. R. and Scudder, G. D. (1990). Sequencing with earliness and tardiness penalties: A review. *Operations Research*, 38(1):22–36.

Cai, X. and Zhou, S. (1999). Stochastic scheduling on parallel machines subject to random breakdowns to minimize expected costs for earliness and tardy jobs. *Operations Research*, 47(3):422–437.

Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932.

Denton, B., Viapiano, J., and Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10(1):13–24.

Eijkemans, M. J. C., Van Houdenhoven, M., Nguyen, T., Boersma, E., Steyerberg, E. W., and Kazemier, G. (2010). Predicting the unpredictable: A new prediction model for operating room times using individual characteristics and the surgeon's estimate. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 112(1):41–49.

Ferrand, Y. B., Magazine, M. J., and Rao, U. S. (2014). Managing operating room efficiency and responsiveness for emergency and elective surgeries: A literature survey. *IIE Transactions on Healthcare Systems Engineering*, 4(1):49–64.

Garey, M. R. and Johnson, D. S. (1978). 'Strong' NP-completeness results: Motivation, examples, and implications. *Journal of the Association for Computing Machinery*, 25(3):499–508.

Guda, H., Dawande, M., Janakiraman, G., and Jung, K. S. (2016). Optimal policy for a stochastic scheduling problem with applications to surgical scheduling. *Production and Operations Management*, 25(7):1194–1202.

Hall, N. G., Kubiak, W., and Sethi, S. P. (1991). Earliness-tardiness scheduling problems, II: Deviation of completion times about a restrictive common due date. *Operations Research*, 39(5):847–856.

Hall, N. G. and Posner, M. E. (1991). Earliness-tardiness scheduling problems, I: Weighted deviation of completion times about a common due date. *Operations Research*, 39(5):836–846.

Hans, E., Wullink, G., Van Houdenhoven, M., and Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, 185(3):1038–1050.

Hulshof, P. J. H., Kortbeek, N., Boucherie, R. J., Hans, E. W., and Bakker, P. J. M. (2012). Taxonomic classification of planning decisions in health care: A structured review of the state of the art in OR/MS. *Health Systems*, 1(2):129–175.

Leeftink, G. and Hans, E. W. (2018). Case mix classification and a benchmark set for surgery scheduling. *Journal of Scheduling*, 21(1):17–33.

Lenstra, J. K., Rinnooy Kan, A. H. G., and Brucker, P. (1977). Complexity of machine scheduling problems. 1(Supplement C):343–362.

Marcon, E., Kharraja, S., and Simonnet, G. (2003). The operating theatre planning by the follow-up of the risk of no realization. *International Journal of Production Economics*, 85(1):83–90.

May, J. H., Strum, D. P., and Vargas, L. G. (2000). Fitting the lognormal distribution to surgical procedure times. *Decision Sciences*, 31(1):129–148.

Mehta, N. B., Wu, J., Molisch, A. F., and Zhang, J. (2007). Approximating a sum of random variables with a lognormal. *IEEE Transactions on Wireless Communications*, 6(7):2690–2699.

Otten, J. W. M., Bos, J., Braaksma, A., and Boucherie, R. J. (2019). Robust surgery scheduling: A model-based overview. *Submitted*.

Pinedo, M. L. (2012). *Scheduling: Theory, Algorithms and Systems*. Springer, New York, USA.

Portougal, V. and Trietsch, D. (2006). Setting due dates in a stochastic single machine environment. *Computers & Operations Research*, 33(6):1681–1694.

Soroush, H. M. (1999). Sequencing and due-date determination in the stochastic single machine problem with earliness and tardiness costs. *European Journal of Operational Research*, 113(2):450–468.

Soroush, H. M. and Fredendall, L. D. (1994). The stochastic single machine scheduling problem with earliness and tardiness costs. *European Journal of Operational Research*, 77(2):287–302.

Van Houdenhoven, M., Van Oostrum, J. M., Hans, E. W., Wullink, G., and Kazemier, G. (2013). Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling. *Anesthesia and Analgesia*, 105(3):707–714.

Van Riet, C. and Demeulemeester, E. (2015). Trade-offs in operating room planning for electives and emergencies: A review. *Operations Research for Health Care*, 7:52–69.

Wachtel, R. E. and Dexter, F. (2009). Influence of the operating room schedule on tardiness from scheduled start times. *Anasthesia and Analgesia*, 108(6):1889–1901.