Proceedings of the 7th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2012)
H. Yang, D. Zeng, O. E. Kundakcioglu, eds.

Optimal Patient-specific Post-operative Surveillance for Vascular Surgery

Shanshan Zhang[1], **Pranav Hanagal**[1], **Peter I. Frazier**[1], **Andrew J. Meltzer**[2], **Darren B. Schneider**[2]

[1] School of Operations Research & Information Engineering, Cornell University
[2] Weill Cornell Medical College

Abstract

We optimize post-operative surveillance schedules for patients who have undergone surgery to restore peripheral blood flow. Post-operative surveillance is performed using periodic duplex ultrasound tests to detect loss of blood flow before it poses a serious risk to the patient. Currently, tests are performed according to the same schedule for all patients, even though problems tend to occur early in some patient groups, and later in others. Moreover, the current schedule is not optimized to when failures occur. We provide a statistical procedure for estimating, from censored observations, the joint distribution of the time at which a problem can be detected in a follow-up visit, and the possibly later time at which the problem threatens the health of the patient. We then find optimal patient-specific follow-up schedules based on this estimation procedure, and compare its performance to the schedule currently in use.

Keywords: survival analysis, vascular surgery, post-operative surveillance

Introduction

We consider patients who have had surgery to repair a loss of blood flow through a limb. Following surgery, patients are monitored in periodic office visits, in which a duplex ultrasound is performed to test whether the repair is continuing to function, i.e, is *patent*, or whether patency has been lost, preventing blood from circulating through the affected limb. If patency loss is detected early enough, then action can be taken to remedy the problem. If it fails to be detected, serious side effects, including the loss of the limb, can result.

Schedules are usually chosen *ad hoc*, without supporting quantitative analysis. For example, current institutional practice at Weill Cornell Vascular and Endovascular Surgery is to schedule follow-up duplex ultrasounds at months 1, 6, 12, 24, 36, and 48 following surgery. This choice of schedule is based on the fact that early failures are more common, but has also been influenced by the convenience of 6 month intervals. Moreover, the same schedule is used for all patients, even though some patients can be identified as being at higher risk for patency loss, and patient-specific covariates can be used to predict when patency-loss will occur [1].

In this paper, we design optimal schedules, which promise both to reduce the number of patients for whom patency loss goes undetected, and to reduce the number of follow-up visits that low-risk patients must undergo. This work has three distinct contributions. First, we present a statistical method for estimating the joint distribution of when patency loss occurs and when it becomes symptomatic. This estimation is challenging because observations of these times are censored. Second, we develop an optimization algorithm for obtaining the optimal schedule given estimates of the parameters of the statistical model. Third, we apply these methods to data to obtain optimal schedules for two different patient groups, and find that the percent of patients with undetected patency loss can be reduced from 59.3% to 56.0% in one group of patients (critical limb ischemia with rest pain only), and from 77.8% to 72.6% in another (critical limb ischemia with tissue loss, which is a more severe form of disease).

Much of the previous work on related problems comes from the reliability engineering literature, and focuses on surveillance of mechanical systems with applications in manufacturing. [4] and [12] introduced the problem of inspection for deteriorating systems, with the objective to minimizing the expected cost of inspection and maintenance. [8] extended the model by allowing the duration of an inspection to be non-negligible. [3] considered the case where failure symptoms are delayed. [9] discussed the system where the inspections are fallible. An overview of this problem may be found in [5, 2, 6]. Performance criteria other than total expected cost have also been studied. For example,

[10] proposed a model to minimize the expected number of inspection times, which aimed to use minimum number of inspection to maintain certain level of availability. [11] minimized the expected delayed time under the constraint of an upper bound for expected cost.

All of this previous work within assumes availability of an infinite number of inspection opportunities, each of which carries a cost which can be expressed in the same units as the cost of a unforeseen failure. In the medical problem that we consider, however, the difficulty of comparing the cost of an extra inspection, or medical test, against the cost of a medical complication arising from an undetected problem lead us to formulate our problem differently: we place a constraint on the number of inspections, and minimize the probability of failing to detect loss of patency before it becomes symptomatic.

A recent paper, [7], considers the problem of scheduling surveillance for patients on the kidney transplant waiting list, to minimize the probability that a patient has undergone an undetected change in medical state when a kidney becomes available for transplantation. This problem is similar to the one we consider here, in that the number of inspections is constrained, and the goal is to minimize the probability that an undetected failure persists at some random checkpoint in time. In [7], this checkpoint is the time at which a kidney becomes available, and is independent of the failure time. In the current paper, however, this checkpoint is the time at which a loss of patency becomes symptomatic, and is highly dependent on when the loss of patency first occurred. This causes differences in both the optimization problem of finding an optimal surveillance schedule, and the statistical estimation techniques required.

We organize our paper in the following way. First, we formulate our optimization problem. Then, we describe the statistical model used in estimating the problem's parameters from data. Then, we introduce optimization algorithms for three different parametric families of distributions used in the problem formulation. Finally, we apply our methods to patient data and obtain optimal schedules for two different patient groups.

Problem Formulation

We model loss of patency (i.e., failure of the original surgical repair) as occurring at a random time τ. Until time τ, any test performed will reveal the repair to be patent in good condition, and on or after τ, a test will reveal that the repair is no longer patent. At time $\tau + \delta$, if the problem has not yet been detected by a scheduled test, the patient will become symptomatic, and will visit the clinic on his or own, suffering a risk of negative side effects from having had the problem go undetected. The density functions of τ and δ are denoted by $f_\tau(t)$ and $f_\delta(t)$ respectively. In this work,we assume τ and δ to be independent.

Our goal is to determine an optimal testing schedule that will minimize the probability that the patient's loss of patency goes undetected and eventually becomes symptomatic, i.e., that the time between patency loss and its detection exceeds δ. Let $t_1 < \cdots < t_m$ be times at which tests will be performed, where t_m is fixed and is the end of the follow-up period under consideration, and the other times $t_1, \ldots, t_{m-1}$ can be chosen to minimize the probability of an undetected failure. Patency is also observed at time $t_0 = 0$, which is the day of surgery.

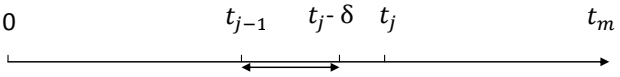

Figure 1: Depicted are four of a sequence of scheduled testing times, $0 = t_0 < t_1 < \cdots < t_m$. Patency loss (failure of the original surgical repair) occurs at a random time τ, and becomes symptomatic at a later time $\tau + \delta$. If, as pictured, patency loss occurs in an interval $[t_{j-1}, t_j - \delta]$ for some j, then it will become symptomatic (at time $\tau + \delta < t_j$) before it is detected. Our goal is to minimize the probability that this occurs.

As illustrated in Figure 1 at right, our goal is equivalent to minimizing the probability that patency loss τ occurs in an interval $(t_{j-1}, t_j - \delta)$ for some j. Thus, our goal is to minimize over $t_1, \ldots, t_{m-1}$ the objective

$$F(t) = \sum_{i=1}^{m} P(t_{i-1} < \tau < t_i - \delta) = \sum_{i=1}^{m} \left(P(\tau > t_{i-1}) \int_0^{t_i - t_{i-1}} f_\delta(v)dv - \int_{v=0}^{t_i - t_{i-1}} f_\delta(v) P(\tau > t_i - v)dv \right)$$
$$\text{s.t.} \quad t_i < t_{i+1} \text{ for } i = 0, \cdots, m-1. \tag{1}$$

Statistical Models

In this section, we introduce the statistical model used to estimate the distributions of τ and δ. First, we define notation. C_i denotes the censor time for patient i (e.g., date of death, or time at which the patient was lost to follow up). s_i denotes the time at which a loss of patency was detected for patient i, whether this was detected in a scheduled test or because it became symptomatic. If no loss of patency is detected, $s_i = C_i$. $f_i(s_i)$ denotes the first scheduled test strictly before s_i. If s_i is itself a scheduled test, then $f(s_i)$ is the preceding test. $u_j (j = 1, \cdots, m)$ is the schedule of times currently in use, and $A_i := \{\, u_j : u_j \in [\tau_i, \tau_i + \delta] \,\}$.

For each patient, we do not observe the times τ and $\tau+\delta$ directly. Instead, we observe only the time at which a loss of patency is detected, whether this occurs in a scheduled test or because it became symptomatic. (We also know the patient's testing schedule.) The way in which loss of patency is detected places an observation of each patient i into one of three categories: $Z_i = 0$ (none), $Z_i = 1$ (clinical), and $Z_i = 2$ (duplex), illustrated in Figure 2. $Z_i = 0$ denotes that patency loss occurs after the censor time C_i. The corresponding likelihood of the observation $(Z_i = 0, s_i)$, as a function of the parameters ν of the distribution P_ν of (τ_i, δ_i) is

$$L_0(\nu) = \prod_{i:Z_i=0} P_\nu(\tau_i \geq f_i(s_i), \tau_i + \delta \geq C_i, C_i = s_i).$$

$Z_i = 1$ denotes that patency loss occurs after one scheduled follow-up time, and the likelihood function is:

$$L_1(\nu) = \prod_{i:Z_i=1} P_\nu(\tau_i+\delta = s_i, \tau_i \geq f(s_i), \tau_i+\delta \leq C_i).$$

$Z_i = 2$ denotes patency loss occurs after one scheduled follow-up time but is detected after the next scheduled follow-up time, and the likelihood function is:

$$L_2(\nu) = \prod_{i:Z_i=2} P_\nu(\tau_i \leq s_i, \tau_i + \delta \geq s_i, s_i = \min(A)).$$

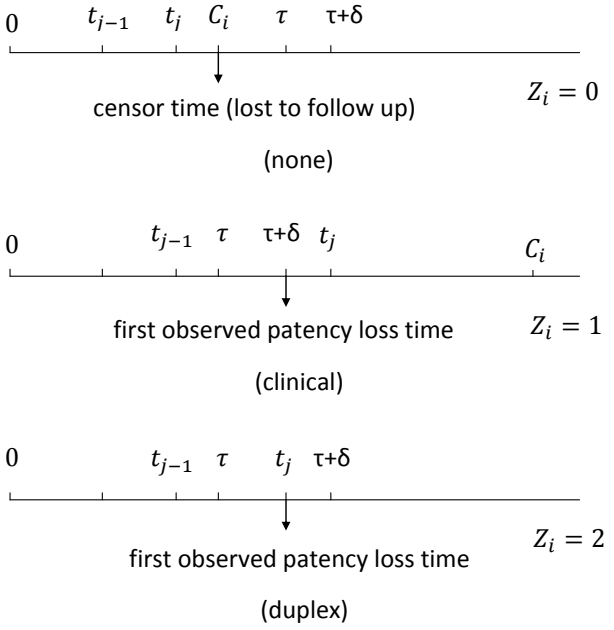

Figure 2: Depicted are three types of patient observations. The first, $Z_i = 0$ or "none", occurs when no failure is observed. The second, $Z_i = 1$ or "clinical", occurs when loss of patency is undetected in a scheduled test and becomes symptomatic. The third, $Z_i = 2$ or "duplex", occurs when loss of patency is observed in a scheduled test. Each type censors the times τ and $\tau + \delta$ differently.

For analytic tractability, we only consider the following distributions for τ and δ: (Case 1) τ follows an Erlang distribution and δ is fixed; (Case 2) both τ and δ follow an exponential distribution with parameters λ and α respectively; (Case 3) τ follows an Erlang distribution with parameters (k, λ) and δ follows an exponential distribution with parameter α. In these three cases, the likelihoods $L_0(\nu), L_1(\nu), L_2(\nu)$ can be competed analytically. Since space is limited, we do not include these calculation, but instead refer to a full version of this paper [13] available online. We then use maximum likelihood estimation, implemented using `fmincon` in Matlab with analytically computed first and second order derivatives, to estimate parameters for each of these models.

Optimization Solutions

In this section, we prove some properties of the optimal solution for the optimization problem, which we then use to develop algorithms for obtaining the optimal follow-up schedule.

Proposition 1. *Suppose $\tau \geq 0$ and $\delta \geq 0$ both satisfy $P(\tau \in B) > 0$ and $P(\delta \in B) > 0$ for all $B \subset [0, t_m]$ with positive Lebesgue measure. If $\{t_i^*\}$ is an optimal solution to (1), then $\{t_i^*\}$ satisfies $t_0 < t_1^* < \cdots < t_{m-1}^* < t_m$. Furthermore, it also satisfies the first order necessary optimality condition*

$$\frac{\partial F(t^*)}{\partial t_i} = \int_0^{t_i - t_{i-1}} f_\delta(v) f_\tau(t_i - v) dv - f_\tau(t_i) \int_0^{t_{i+1}-t_i} f_\delta(v) dv = 0 \;\text{ for } 0 < i < m.$$

The above proposition indicates that the optimal follow-up times are determined by the first follow-up time. In other words, once the first optimal follow up time is determined, we can obtain other optimal follow up times recursively by using the first order necessary optimality condition. Moreover, it also implies that an optimal solution exists, since we minimize a continuous function over a bounded set and all the optimal solutions lie in the interior of the feasible set.

Case 1: δ is fixed and τ follows a Weibull distribution. To gain insight into the optimization problem, we first consider a very special case when δ is fixed. For this case, we can solve the optimization problem analytically. We assume patency loss τ follows a Weibull distribution with parameters (k, λ) and density function $S(t) = e^{-(t/\lambda)^k}$. The natural constraints for follow-up times should be $t_i - \delta \geq t_{i-1}$ for $i = 1, \cdots, m$. The reason is that if we detect patency loss at t_i, then the patient will not be symptomatic before $t_i + \delta$. So it is reasonable to schedule the next follow up time after $t_i + \delta$. Then the optimization can be written as follows:

$$\begin{array}{l} \sum_{i=1}^{m} [e^{-(t_{i-1}/\lambda)^k} - e^{-((t_i-\delta)/\lambda)^k}] \\ \text{s.t. } t_i - \delta \geq t_{i-1} \text{ for } i = 1, 2, 3, ...m. \end{array}$$

Proposition 2. *Let $H(t) = e^{-(t/\lambda)^k} - e^{-((t-\delta)/\lambda)^k}$. We assume that $H(t)$ either is increasing or has a unique minimizer x^* with $x^* + (m-1)\delta \leq t_m$ on interval $[\delta, t_m]$. Suppose $t^* = (t_1^*, \cdots, t_{m-1}^*)$ is an optimal solution. Then t^* should have the following structure*

$$t^* = (t_1^*, t_1^* + \delta, t_1^* + (m-2)\delta).$$

Furthermore, all optimal follow up times are determined by the optimal first follow up time t_1^*. Therefore, we can formulate the optimization problem as

$$\min \sum_{i=1}^{m-1} H(x + (i-1)\delta) \text{ s.t. } \delta \leq x \leq t_m - (m-1)\delta.$$

Case 2: τ and δ follow exponential distributions. If we assume that τ and δ follow exponential distributions with parameters λ and α respectively, the corresponding optimization problem is:

$$\begin{array}{ll} \min : F(t) = \sum_{i=1}^{m} \left(e^{-\lambda t_{i-1}} \int_0^{t_i - t_{i-1}} \alpha e^{-\alpha v} dv - e^{-\lambda t_i} \int_0^{t_i - t_{i-1}} \alpha e^{-(\alpha-\lambda)v} dv \right) & (2) \\ \text{s.t.} \quad t_i < t_{i+1} \quad \text{for} \quad i = 0, \cdots, m-1. & \end{array}$$

Then, according to Proposition 1, we have the first order necessary optimal condition for (2):

$$\int_0^{t_i - t_{i-1}} e^{-(\alpha-\lambda)v} dv = \int_0^{t_{i+1} - t_i} e^{-\alpha v} dv. \quad (3)$$

Proposition 3. *Suppose both τ and δ have exponential distributions with parameters λ and τ respectively. Then there exists a unique $(t_1, \cdots, t_{m-1})$ satisfies the first order necessary optimal condition, i.e. $\frac{\partial F(t)}{\partial t_i} = 0$, which is also the optimal solution for (2).*

We design an optimization algorithm to obtain a solution which satisfies the first order optimality condition, which is also the optimal solution by Proposition 3. The optimization problem uses the idea of bisection algorithm to solve $m-1$ equations. The algorithm is included in [13].

Case 3: τ follows an Erlang distribution and δ follows an exponential distribution. If we assume that τ follows an Erlang distribution with parameters (k, λ) and δ follows an exponential distribution with parameter α, the corresponding optimization problem is as follows:

$$\begin{array}{ll} \min : F(t) = \sum_{i=1}^{m} \left[\sum_{n=0}^{k-1} \left(\frac{\lambda^n}{n!} e^{-\lambda t_{i-1}} t_{i-1}^n (1 - e^{-\alpha(t_i - t_{i-1})}) - \frac{\lambda^n}{n!} e^{-\alpha t_i} \int_{v=t_{i-1}}^{t_i} \alpha e^{-(\lambda-\alpha)v} v^n dv \right) \right] & (4) \\ \text{s.t.} \quad t_i < t_{i+1} \quad \text{for} \quad i = 0, \cdots, m-1. & \end{array}$$

According to Proposition 1, we have the first order necessary optimal condition for (4):

$$e^{-\alpha t_i} \int_{t_{i-1}}^{t_i} v^{k-1} e^{-(\lambda-\alpha)v} dv = t_i^{k-1} e^{-\lambda t_i} \int_0^{t_{i+1} - t_i} e^{-\alpha v} dv. \quad (5)$$

Results

We obtained data for 239 patients with critical limb ischemia, and divided them into two groups: those with rest pain only; and those with tissue loss, which is a more severe condition. We modeled each group separately. For each patient, this data contains the first observed patency loss and failure type (none, clinical, duplex). It does not, however, include the last office visit for the patient before the observed failures or censor time, which is necessary in our statistical procedure for fitting the parameters of τ and δ. To overcome this limitation, we assumed that each patient followed the current recommended schedule of $u_1 = 1(\text{month}), u_2 = 6, u_3 = 12, u_4 = 24, u_5 = 36, u_6 = 48$.

We then performed maximum likelihood estimation for the model with τ Erlang and δ exponential across the three parameters k, λ, and α, for the shape parameter $k \in \{1, 2, 3\}$ and $\lambda, \alpha \geq 0$, using Matlab's `fmincon`. We found the best fit at $k = 1$ for both patient groups, which corresponds to an exponential distribution for τ. The fitted parameters were $\lambda = 0.0338, \alpha = 0.4500$ for the rest pain group, and $\lambda = 0.0530, \alpha = 0.9022$ for the tissue loss group. These fitted values show that patency loss occurs more quickly in the tissue loss group, which is consistent with these patients' increased severity of disease. This patient group also has a shorter duration between patency loss and onset of symptoms.

Table 1 shows the results of computing the optimal schedule for the fitted distributions for each of these two patient groups. The percentage of patients with undetected patency loss that eventually becomes symptomatic is computed under the statistical model for both the current schedule (at 1, 6, 12, 24, 36, 48 months) and the optimal one. The optimal schedule is able to significantly reduce the percentage of patients with symptomatic patency.

patient type	optimal schedule (in months)	percentage of patients with symptomatic patency loss	
		under optimal schedule	under current schedule
rest pain	3.3, 7.0, 11.1, 16, 22.2, 48	56.0%	59.3%
tissue loss	1.8, 3.9, 6.1, 8.8, 12.1, 48	72.6%	77.8%

Table 1: Results of analysis: Table shows the optimal schedule (in months since day of surgery) and the percentage of patients for whom patency loss is undetected in a scheduled test and later becomes symptomatic, under both the current schedule, and the optimal one. This is shown for two groups of patients: rest pain only; and tissue loss.

Figure 3 compares the empirical cumulative distribution functions (cdfs) with predictions based on the fitted statistical model. For each of the two patient groups, we calculated the empirical cdf of s_i for duplex failures, and then again for clinical failures. We then computed what these cdfs would be under the fitted statistical model with the current schedule. While the empirical and predicted cdfs are roughly comparable, and some deviation is to be expected due to noise from the finite sample size, we believe that there is room for improvement, and that allowing more general distributions for τ and δ may result in a better fit.

Conclusion

We have described a statistical method for estimating the joint distribution of when patency loss occurs, and when this patency loss becomes symptomatic. This statistical method is able to obtain estimates in the face of a type of censoring specific to this surveillance problem. We then described optimization algorithms able to compute the optimal schedule for three different parametric families for these two times. We then applied these statistical and optimization methods to data, obtaining optimal schedules that improved predicted performance significantly over the existing schedule.

While these results are encouraging, and strongly suggest that substantial improvements can be made by optimizing the schedule to patient groups, the current work has a number of limitations. First, our statistical analysis assumed that patients follow the existing schedule exactly, while data on duplex failures are not necessarily close to scheduled times, demonstrating that patients deviate from the recommended schedule. In future work, we will overcome this limitation by incorporating additional data on when each patient had his or her tests. Second, our statistical analysis used relatively restrictive parametric families of distributions for τ and δ, which in particular do not allow for the decreasing failure rate observed in [1]. This statistical analysis also assumed that τ and δ are independent, while this may not necessarily be true in practice. We believe that these first two limitations are the cause for the poor fit

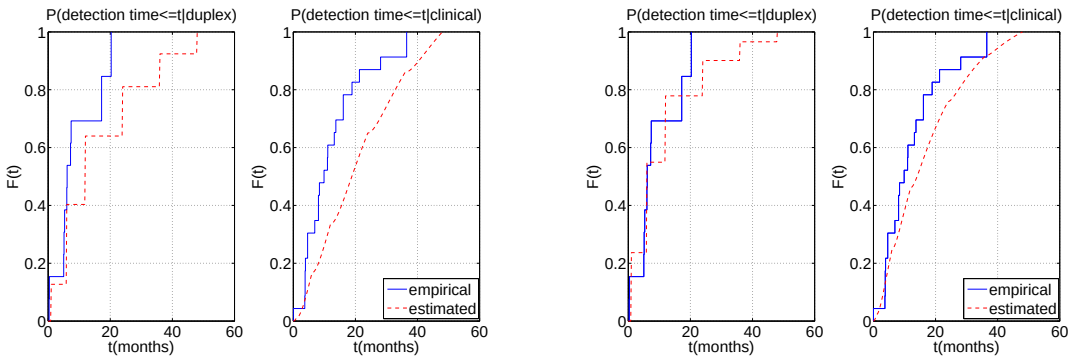

Figure 3: Model validation: Each pair of plots, one for patients with rest pain only (the left pair) and the other for patients with tissue loss (the right pair), shows the conditional distribution of the time that a failure is detected, given the failure type. The left plot in each pair shows this for failures detected in duplex ultrasound, and the right plot for failures detected when they become symptomatic. Each plot shows two curves, one coming from empirical data, while the other is computed using the fitted model and the current schedule.

in Figure 3. Third, our optimization problem could have modeled patients as scheduling tests at random times near the recommended schedule, rather than following the schedule exactly. The optimal schedule under this more robust model would then have an additional robustness property.

References

[1] A. J. Meltzer, A. Graham, P. H. Connolly, J. K. Karwowski, H. L. Bush, P. I. Frazier, and D. B. Schneider, 2012, "Risk Factors for Early Failure after Peripheral Endovascular Intervention: Application of a Reliability Engineering Approach," Ann. Vasc. Surg., in press, available online Sep. 2012.

[2] H. Wang, 2002, "A survery of maintenance policies of deteriorating systems," Eur. J. Oper. Res., 139(3), 469-489.

[3] B. Sengupta, 1980, "Inspection procedures when failure symstoms are delayed," Oper. Res., 28(3), 768-776.

[4] R. E. Barlow, L. C. Hunter, and F. Proschan, 1963, "Optimum checking procedures," J. Soc. Ind. Appl. Math., 11(4), 1078-1095.

[5] Y. S. Sherif and M. L. Smith, 1981, "Optimal maintenance models for systems subject to failure-a review," Nav. Res. Logist. Q., 28(1), 47-74.

[6] A. Chelbi and D. Ait-Kadi, 1999, "An optimal inspection strategy for randomly failing equipment," Reliab. Eng. Syst. Saf., 63(2), 127-131.

[7] A. Sabouri, S. M. Shechter, and W. T. Huh, 2012, "Inspection models with application to screening the kidney transplant waiting list," (manuscript).

[8] H. Luss and Z. Kander, 1974, "Inspection policies when duration of checking is non-negligible," Oper. Res. Q., 25(2), 299-309.

[9] G. Parmigiani, 1996, "Optimal scheduling of fallible inspections," Oper. Res., 44(2), 360-367.

[10] G. Parmigiani, 1994, "Inspection times for stand-by units," J. Appl. Probab., 31(4), 1015-1025.

[11] M. Klein and R. Kirch, 1994, "Surveillance Schedules for Medical Examinations," Manage. Sci., 20(10), 1403-1409.

[12] R. E. Barlow and F. Proschan, 1985, Mathematical Theory of Reliability, Wiley, New York.

[13] S. Zhang, P. Hanagal, P. I. Frazier, A. J. Meltzer, and D. B. Schneider, "Optimal Patient-specific Post-operative Surveillance for Vascular Surgeries", in preparation, available online http://people.orie.cornell.edu/sz254/.