

Scheduling medical procedures: how one single delay begets multiple subsequent delays

AMNON SONNENBERG^{†‡*} and BRADFORD R. CRAIN[¶]

[†]Portland VA Medical Center, Portland, OR, USA

[‡]Department of Medicine, Oregon Health & Science University, Portland, OR, USA

[¶]Department of Mathematics, Portland State University, Portland, OR, USA

Background: Delay is a common feature of medical disease management. Delays occur because schedules are filled, patients forget their appointment, equipment is unavailable, or because medical and non-medical complications interfere with the planned procedure. The aim of the present analysis is to model how one single delay can lead to multiple subsequent delays.

Methods: The consecutive stream of delays is analyzed in terms of a stochastic process comprising of a random sum of random time periods. Any untoward event causes a procedural delay, which provides a time window of opportunity for yet another delaying event to occur.

Results: The stochastic model explains why even a single initial delay can easily lead to a multitude of subsequent delays. The expected overall delay is always longer than the initial delay caused by the deferment of the initial procedure. The analysis demonstrates how in individual patients an initially short delay may subsequently expand into days or weeks.

Conclusion: Because a single delay can easily burgeon into a lengthy series of multiple delays, the primary goal should be to avoid the precipitating delay at the onset.

Keywords: Health service research; Medical decision analysis; Random sums; Stochastic modeling

Introduction

Delays are common and unfortunate features of all medical care. Most delays among hospitalized patients stem from scheduling of diagnostic test procedures [1,2]. The types of events that delay patient management relate to a large variety of medical, organizational, administrative and technical obstacles that often render a timely and expeditious management difficult. Delays occur, for instance, because schedules are already filled and overbooked, patients forget their appointment and need to be rebooked, instruments become broken and unavailable or because comorbid conditions and new complications interfere with the performance of a planned procedure. A delay in establishing a diagnosis or initiating treatment can result in severe medical consequences [3–5]. Frequently, situations arise when one single delay leads to multiple consecutive delays. It then becomes a rather frustrating experience for the patient, as well as the physician, to appreciate the urgency of a particular medical procedure, yet be unable to get such procedure done. Although, the procedure may be absolutely indicated for diagnostic confirmation or therapeutic resolution, nevertheless, the clinician may find it

extremely difficult to reach its elusive goals because of the many ensuing delays. It appears as if an initial delay in performing the procedure precipitates the occurrence of various other medical and non-medical events that push the planned procedure further and further away.

As a typical scenario, consider a 79-year-old man who was admitted to the hospital for post-prandial abdominal pain. His serum level of total bilirubin measured 4.8 mg/dl and his alkaline phosphatase measured 204 U/l. An abdominal ultrasound showed multiple stones in the gallbladder and a dilated common bile duct of 12 mm diameter. The patient had been treated with low-dose aspirin for cardiovascular prevention and coumadin for atrial fibrillation. When first seen by the Gastroenterology Consult Service on Tuesday, it was recommended to take the patient off his anticoagulative medication and schedule an endoscopic retrograde cholangiography (ERC) with possible stone extraction from his common bile duct on Thursday. When tested on Tuesday morning, the patient's INR value of 2.2 was considered still too high and the procedure was postponed until Monday. On Friday, however, the ERC endoscope broke down during a prior procedure on another patient and with the second back-up instrument still in repair, no other instrument was

*Corresponding author. Email: sonnenbe@ohsu.edu

available to do the scheduled ERC on Monday. Because the two physicians performing ERC left to attend a two-day conference, the next available time slot for ERC was only on the following Friday. In the meantime, the patients elevated serum bilirubin and alkaline phosphatase trended down and it was hypothesized that the initial cholestasis had stemmed from passed gallstone. Rather than wait for the ERC, the patient underwent an uneventful laparoscopic cholecystectomy. An intra-operative cholangiogram again revealed a dilated common bile duct with a suspected gallstone lodged above the papilla. The ERC was eventually done on the following Friday, that is, seventeen days after the initial hospital admission and fifteen later than first planned.

The scenario from above represents only one example of many similar clinical instances, where the cumulative occurrence of several unpredicted events can result in an extensive overall delay. Why does this happen? The aim of the present article is to describe this process and analyze the mechanisms underlying its occurrence. The analysis is not concerned with waiting queues in general or patient flow through medical systems, but focused solely on how one delaying event prepares the ground for additional subsequent delays. A stochastic model is developed to provide estimates for expected delays in a large set of potential clinical scenarios.

Methods

In the present context, the term “event” refers to an incident which causes delay. The term “delay” refers to the excess time needed to accomplish a medical task. Rarely does one delaying event come alone, as frequently the first event lends to the occurrence of a second, third, etc. subsequent delaying event. The initial delay caused by the first event opens up a time window, during which another delaying event can occur. Let d be the average length of time of a delay. During the time window of length d , the probability for a second delay arises. The second delay provides a new time window for yet another delaying event. A delaying event can occur anytime within the window of opportunity provided by the previous delay. Because on the average, the next event will occur in the middle of the previous delay, each new event adds $d/2$ to the overall delay. The overall delay D equals

$$D = d_1 + d_2 + d_3 + \dots + d_N, \tag{1}$$

where the individual delays d_i are assumed to be independent and identically and uniformly distributed over an interval $[0,d]$ with an expected value $E[d_i] = d/2$ and a variance $\text{var}[d_i] = d^2/12$ [6]. Once the first initiating delay has occurred, the probability for the occurrence of any secondary delaying event is p and for its non-occurrence $q = 1 - p$. The aim of the following analysis is to develop an estimate for the expected length of the overall delay and its variance. The number of secondary delaying events N is random. After the primary delay has

occurred, the probability for none, one, two, etc. secondary delaying events are: $P(N = 0) = q$, $P(N = 1) = qp$, $P(N = 2) = qp^2$, etc. with the corresponding geometric probability mass function (pmf):

$$\text{pmf} = \sum_{n=0}^{\infty} p^n q. \tag{2}$$

The expected overall number of secondary events $E[N]$ is calculated as:

$$\begin{aligned} E[N] &= \sum_{n=0}^{\infty} np^n q = pq \sum_{n=1}^{\infty} np^{n-1} = pq \frac{1}{(1-p)^2} \\ &= \frac{p}{q}. \end{aligned} \tag{3}$$

$$\begin{aligned} E[N(N-1)] &= \sum_{n=0}^{\infty} n(n-1)p^n q \\ &= p^2 q \sum_{n=2}^{\infty} n(n-1)p^{n-2} \\ &= p^2 q \frac{2}{(1-p)^3} = \frac{2p^2}{q^2}. \end{aligned} \tag{4}$$

Since in general, $\text{var}[N] = E[N^2] - (E[N])^2$, the two formulas from above yield:

$$\begin{aligned} \text{var}[N] &= E[N(N-1) + N] - (E[N])^2 \\ &= \frac{2p^2}{q^2} + \frac{p}{q} - \frac{p^2}{q^2} = \frac{p}{q^2}. \end{aligned} \tag{5}$$

The expected overall delay $E[D]$ corresponds to the expected number of delays $E[N]$ multiplied by the expected length of the individual delays $E[d_i]$. The last delay, running its course uninterrupted by any subsequent event, is expected to be twice as long as the preceding delays, that is $E[d_{\text{last}}] = 2E[d_i] = 2d/2 = d$. Hence:

$$E[D] = d + E[N]E[d_i] = d + \frac{pd}{q2}. \tag{6}$$

The variance of the overall delay $\text{var}[D]$ is calculated according to the general formula for the variance of a sum of random variables [7,8]:

$$\begin{aligned} \text{var}[D] &= \text{var}[d_1 + d_2 + d_3 + \dots + d_N] = \text{var}[Nd_i] \\ &= (E[d_i])^2 \text{var}[N] + E[N] \text{var}[d_i]. \end{aligned} \tag{7}$$

Substituting with the terms from above, equation (7) changes to:

$$\text{var}[D] = \frac{d^2 p}{4 q^2} + \frac{pd^2}{q12}, \quad \text{with SD} = \sqrt{\text{var}[D]}. \tag{8}$$

The stochastic model of consecutive delays was simulated on a computer. A good agreement

(with errors of less than 5%) was obtained between the predicted and the simulated values of the overall delay D and its standard deviation SD.

Results

Using equation (6) given in the methods, an event probability of $p = 50\%$ and a delay of $d = 3$ days result in an overall expected delay of $D = 4.5$ days. Similarly, a higher event probability of $p = 90\%$ results in an overall expected delay of $D = 16.5$ days. As the event probability p increases, the overall delay D becomes longer and longer. With a probability $p \approx 1$ close to certainty that each new delay will provide sufficient time for yet another event to occur and cause another delay, the overall length of D stretches out to infinity. Figure 1 shows the overall delay as function of delay probability and length.

Not every patient, however, will necessarily experience the overall delay D . First, the chain of events delaying the procedure may stop at any given point in time, when due to their random nature, no new events occur. Second, since the lengths of individual delays fluctuate around an average value d , some events may cause much shorter or much longer delays than d . The standard deviation of the overall delay is given by equation (8) in the methods, its result being easily calculated on a spreadsheet or handheld calculator. In the first example from above, $p = 50\%$, $d = 3$ days, $D = 4.5$ days and the standard deviation of the expected delay is calculated as $SD = 2.3$ days. The 95% confidence interval of the expected delay ranges between $D \pm 1.96 SD$, that is, from 0 to 9.0 days. In the second example, $p = 90\%$, $d = 3$ days and $D = 16.5$ days. The standard deviation of the expected delay is $SD = 14.5$ days. Hence, the 95% confidence interval for the expected delay of 16.5 days ranges between $16.5 \pm 1.96 \times 14.5$ days, that is, from 0 to 44.9 days.

Figure 2 serves as a general guide to estimate the expected overall delay for a large set of possible scenarios. In the four graphs, the probability values of delays are varied between 0 and 100%, while the average lengths of the individual delays are varied between 1, 3, 5 and 10 days. Shorter delays may occur, for instance, when key medical personnel are not available, instruments break down or abnormal laboratory values are encountered. Mid-sized delays are frequently associated with patients harboring comorbid conditions or with organizational obstacles that limit the number of time slots available for procedures. Long delays are most likely encountered in instances of complex medical procedures or surgical interventions that are difficult to schedule and that require sophisticated equipment or the interplay of multiple medical subspecialties. As a general rule, all types of delays tend to be longer in the elderly as opposed to younger patients. Although, the initial delay stems from deferment of a medical procedure, subsequent delays may be caused occasionally by events outside medicine and unrelated to the underlying disease. In each graph

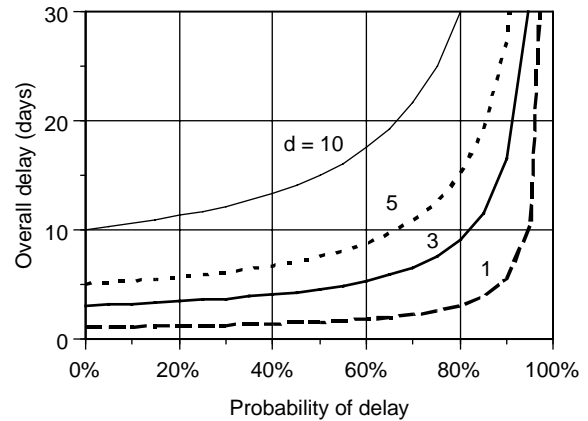


Figure 1. Expected overall delay as function of probability and lengths (d) of recurrent individual delays.

of figure 2, the gray area represents the confidence interval between the mean overall delay plus two standard deviations. As the two upper graphs demonstrate, even a short initial delay of 1–3 days can easily expand into a protracted overall delay, especially, if the tendency for repeat delays creeps above 50%. The lower two graphs demonstrate that, in dealing with long individual delays, the expected overall delay is markedly increased even with low underlying risks of delay. As a general rule, the overall delay must always be expected to turn out longer than the initial precipitating delay.

Discussion

Sometimes, it can become a rather frustrating experience for a physician managing patients to pursue an obvious diagnostic or therapeutic goal, but then encounter a seemingly endless number of obstacles that push a planned intervention further away into the future and keep the physician from reaching his/her goal. There seems to be a continued interference by a slew of minor and often trivial events that sidetrack the entire work-up, complicate the medical pursuit and lead down a lengthy and convoluted path, before the medical goal that has been so clearly discernible from the onset is finally achieved. Such delays are a common feature among hospitalized patients and are similarly encountered in the health care systems from different countries [9–12].

In the present analysis, a model is developed to describe the stochastic process that underlies the recurrence of medical and non-medical events delaying medical procedures. In essence, any untoward event is modeled to cause a procedural delay, which provides a time window of opportunity for yet another delaying event to occur. The consecutive stream of delays is thus analyzed in terms of a random sum of random time periods. In this model, independent events are characterized by a similar rate of occurrence and a similar average length of delay. As in any mathematical model, these assumptions simplify

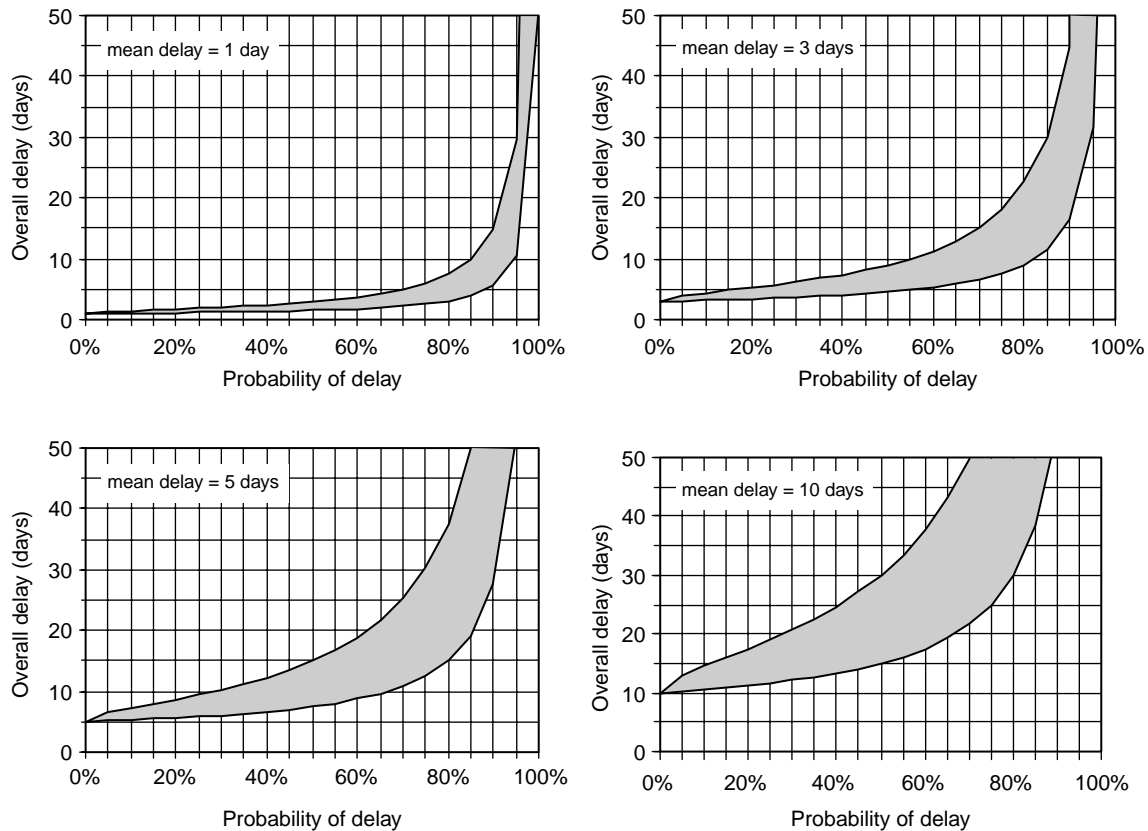


Figure 2. Expected length of the overall delay in relation to the probability of recurrent individual delays. In the four graphs, the mean lengths of the individual delays are varied between 1, 3, 5 and 10 days. The gray area represents the confidence interval between the mean overall delay plus two standard deviations.

the reality of clinical medicine and they do not always represent the entire complexity encountered in the management of actual patients. To derive the estimates from above, several simplifying assumptions have to be made. The delays are assumed to be independent of each other and equally uniformly distributed. However, actual delays in clinical practice may be distributed according to a lognormal, exponential or some other statistical distribution. If two delays occur consecutively, the preceding first delay is assumed to become superseded by the subsequent second delay. In reality, some delays could be additive. Consecutive delaying events may be linked causally and not only temporally. The lengths of consecutive delays may depend on each other or prolong over time as they accumulate in the individual patient. In spite of its simplifications, however, the model provides useful insights into the occurrence of delays associated with medical procedures. The stochastic model explains how even a short initial delay can easily result in a lengthy overall delay. It also demonstrates why in some patients the overall delay may extend over days or even weeks.

There are other aspects of delays that have not been dealt with by this analysis and it should be stressed that the present model was not intended as a general analysis of delays in healthcare. Delay is a common feature of medical practice and occurs in a variety of instances and forms, for example, delays before admission to the

healthcare system, delays in scheduling and appointments, time spent in awaiting areas to be seen by a physician, nurse or technician and waiting times spent before surgery, X-ray or other medical procedures. A large body of stochastic models and administrative instruments has been developed to analyze and manage patient flow within medical systems. Queuing theory has been utilized to estimate lengths of waiting lines and waiting times in patient scheduling and hospital operations [13–16]. The instruments of theory of constraints have been used to identify components of delays and how to eliminate them in order to improve system performance [17–19]. In contrast with such administrative goals to streamline an overburdened medical system, the present model was aimed to illustrate how the management of an individual patient becomes derailed by a series of short delays that can accumulate into one major delay. The analysis has been focused on the perspective of a physician struggling to expedite the patient's work-up rather than the perspective of an administrator trying to improve the hospital's throughput.

What are the benefits of such analysis and what conclusion can be drawn from it? First and foremost, it is important for a practicing physician to be aware of the fact that one delaying event may not come alone and that even a short delay can readily lead to a cascade of additional delaying events. Although, removing or shortening the

initial delay does not guarantee that further downstream delays will not occur, it reduces the general risk for long delays. The stochastic model helps unravel the underlying mechanisms and clarify the seemingly mystifying occurrence of cascading delays. Obviously, no dark clouds hovering over the patient's head or any bad luck are involved but rather the laws of probability.

Because a single delay can easily burgeon into a lengthy series of multiple delays, the physician's primary goal should be to avoid the precipitating delay at the onset. As a simple measure to prevent delays, patients should be discharged from a medical system as quickly as possible to diminish their exposure to the whims of such systems. By staying longer within a given medical system, the patient continues to be vulnerable to its potential risks and failures. In general, patients with *a priori* high proclivity for any delaying events are also more susceptible to protracted delays. Elderly patients, for instance, are more prone to medical and non-medical complications and they require longer time periods to recover from untoward events [20–22]. It would behoove the physician, therefore, to be especially wary of delays in elderly patients and to see such patients undergo an expeditious work-up and become discharged from the hospital early. Occasionally, the only means to interrupt a chain of recurrent delays consists of pulling a patient out of the system, forgo the procedure for the time being and start from scratch altogether by scheduling the procedure for a much later date.

References

- [1] Schlupe, M., Bogousslavsky, J., Regli, F., Tendon, M., Prod'hom, L.S. and Kleiber, C., 1994, Justification of hospital days and epidemiology of discharge delays in a department of neurology. *Neuroepidemiology*, **13**, 40–49.
- [2] Selker, H.P., Beshansky, J.R., Pauker, S.G. and Kassirer, J.P., 1989, The epidemiology of delays in a teaching hospital. The development and use of a tool that detects unnecessary hospital days. *Medical Care*, **27**, 112–129.
- [3] Funch, D.P., 1985, Diagnostic delay in symptomatic colorectal cancer. *Cancer*, **56**, 2120–2124.
- [4] Iregui, M., Ward, S., Sherman, G., Fraser, V.J. and Kollef, M.H., 2002, Clinical importance of delays in the initiation of appropriate antibiotic treatment for ventilator-associated pneumonia. *Chest*, **122**, 262–268.
- [5] Von Titte, S.N., McCabe, C.J. and Ottinger, L.W., 1996, Delayed appendectomy for appendicitis: causes and consequences. *American Journal of Emergency Medicine*, **14**, 620–662.
- [6] Parzen, E., 1960, *Modern Probability Theory and its Application* (New York: John Wiley and Sons), p. 210.
- [7] Drake, A.W., 1967, *Fundamentals of Applied Probability Theory* (New York: McGraw-Hill), pp. 108–112.
- [8] Taylor, H.M. and Karlin, S., 1998, *An Introduction Stochastic Modeling*, 3rd ed (New York: Academic Press), pp. 70–79.
- [9] Chiu, H.C., Lee, L.J., Hsieh, H.M. and Mau, L.W., 2003, Inappropriate hospital utilization for long-stay patients in southern Taiwan. *Kaohsiung Journal of Medical Sciences*, **19**, 225–232.
- [10] Fellin, G., Apolone, G., Tampieri, A., Bevilacqua, L., Meregalli, G., Minella, C. and Liberati, A., 1995, Appropriateness of hospital use: an overview of Italian studies. *International Journal for Quality in Health Care*, **7**, 219–225.
- [11] Panis, L.J., Verheggen, F.W. and Pop, P., 2002, To stay or not to stay. The assessment of appropriate hospital stay: a Dutch report. *International Journal for Quality in Health Care*, **14**, 55–67.
- [12] Schiff, E., Modan, B., Barzilay, Z., Blumstein, Z., Fuchs, Z. and Mozes, B., 1993, Patterns of unjustified pediatric hospital stay. *Israel Journal of Medical Sciences*, **29**, 33–36.
- [13] George, J.A., Fox, D.R. and Canvin, R.W., 1983, A hospital throughput model in the context of long waiting lists. *Journal of the Operational Research Society*, **34**, 27–35.
- [14] Gotein, M., 1990, Waiting patiently. *New England Journal of Medicine*, **323**, 604–608.
- [15] Sonnenberg, A., 2000, Waiting lines in the endoscopy unit. *Gastrointestinal Endoscopy*, **52**, 517–524.
- [16] Worthington, D.J., 1987, Queueing models for hospital waiting lists. *Journal of the Operational Research Society*, **38**, 413–422.
- [17] Breen, A.M., Burton-Houle, T. and Aron, D.C., 2002, Applying the theory of constraints in health care: Part 1—the philosophy. *Quality Management in Health Care*, **10**, 40–46.
- [18] Haraden, C., Nolan, T., Resar, R., Litvak, E. and Members of IHI's IMPACT network and pursuing perfection initiatives, 2003, *Optimizing patient flow. Moving patients smoothly through acute care settings* (Cambridge, MA: Institute for Healthcare Improvement). Available at: <http://www.ihl.org>.
- [19] McNutt, R.A., Abrams, R. and Aron, D.C., for the Patient Safety Committee, 2002, Patient safety efforts should focus on medical errors. *Journal of the American Medical Association*, **287**, 1997–2001.
- [20] Panis, L.J., Gooskens, M., Verheggen, F.W., Pop, P. and Prins, M.H., 2003, Predictors of inappropriate hospital stay: a clinical case study. *International Journal for Quality in Health Care*, **15**, 57–65.
- [21] Rockwood, K., 1990, Delays in the discharge of elderly patients. *Journal of Clinical Epidemiology*, **43**, 971–975.
- [22] Victor, C.R., Healy, J., Thomas, A. and Seargeant, J., 2000, Older patients and delayed discharge from hospital. *Health and Social Care in the Community*, **8**, 443–452.