2015 IEEE International Conference on
Automation Science and Engineering (CASE)
Aug 24-28, 2015. Gothenburg, Sweden

Analysis of Multi-Patient Rapid Response Processes: An Iterative Approach

Xiaolei Xie, Zexian Zeng, Jingshan Li, Colleen H. Swartz and Paul DePriest

***Abstract*— In acute care, providers need to response quickly to patients deterioration. However, a physician's availability can be limited if multiple patients are declining simultaneously. To study the multi-patient rapid response process, a complex network model with split, merge and parallel structures is introduced, and iteration procedures are presented to evaluate system performance. It is shown that such procedures are convergent and lead to accurate performance evaluation.**

Keywords: Rapid response, decision time, iterations, patient deterioration, provider availability

I. Introduction

Rapid response teams (RRTs) have been introduced in many hospitals to quickly evaluate, triage, and treat deteriorating patients [1], after the release of Institute of Medicine's report "To Err is Human" [2]. However, there is no strong evidence that the implementation of RRTs has achieved the goal of reducing the frequency and severity of negative outcomes [3]. The rapid response process involves collaborative and integrated operations of multiple care providers, such as nurse, RRT, intern, resident, fellow, and attending, from different departments (e.g., floor, ICU). A systematic study from the point of view of a rapid response system (RRS) is necessary [4]. Developing an analytical model to provide a quantitative perspective of RRS is desirable.

Although rapid response has been addressed extensively in clinical studies [5], [6], [7], the investigation using a systems engineering approach is quite limited. In papers [8], [9], an analytical framework has been proposed to study rapid response system with a single patient. The decision time, i.e., the time from deterioration detection by the nurse to a clinical treatment decision being made by the physician, and its variabilities have been derived. The response time performance, i.e., the probability that the decision is made within a given time interval, has been introduced and evaluated. In addition, improvement efforts by identifying the most critical response process that impedes rapid decision making, referred to as the bottleneck, have been carried out. Bottleneck indicators based on the data collected on the hospital floor have been introduced. In practice, the care providers typically take care of multiple patients, who may deteriorate simultaneously. Thus, their availability is limited, which may result in waiting time and delayed decisions. How to analyze such scenarios to provide an estimate of the overall decision time is still unknown. This paper intends to contribute to this end.

The main contribution of this paper is in developing an iterative method to evaluate the performance of multi-patient rapid response process. By taking into account the waiting time due to limited availability of providers through recursive procedures, the mean decision time can be estimated.

The remainder of the paper is structured as follows: Section II describes the care delivery process in multi-patient rapid response system and formulates the problem. Section III introduces a two-level resource iteration method to estimate mean decision time. Section IV discusses the accuracy of the method. Finally, conclusions are given in Section V. Due to page limitation, all proofs are omitted and can be found in [10].

II. System Description and Problem Formulation

Consider the RRS operations under a multi-layer referring mechanism shown in Figure 1. The nurse will initiate rapid response after detection of a patient's declining through vital signs. He/she may call the RRT, the intern or the resident, or call the RRT and another provider jointly (either intern, or resident, or fellow, or attending). If the provider can make a decision, appropriate treatment can be carried out. Otherwise, higher level help will be requested (e.g., intern to resident, resident to fellow, and fellow to attending). If the attending is informed, he/she will make a final decision. The decisions include one of the following: ICU, step down, telemetry, or stay, where "ICU" implies admission to intensive care unit, "step down" represents progressive care – a level lower than ICU, "tele" (telemetry) refers to moving from acute care to a monitored bed (where physiologic monitor presents) – another level lower than step down, and "stay" stands for continuing observation. (Note that the RRT can only make a "stay" decision.)

Define a resource set of intern, RRT, resident, fellow and attending, denoted as $X_1 = \{int, rrt, res, f, a\}$, respectively. The joint group by RRT and another provider is denoted as set $X_2 = \{rrt\&int, rrt\&res, rrt\&f, rrt\&a\}$. In addition, let t_d be the decision time (from declining to final decision making) and T_d as its mean.

When multiple patients are declining, the provider can only take care of one patient each time. Thus, other patients may need to wait. The response time of each provider follows a general distribution. The waiting time is a function of all

This work is supported in part by NSF Grants No. CMMI-1233807 and 1234636 and the BeiJing Talents Fund Grant 2014000020124G118 of China.

X. Xie is with Department of Industrial Engineering, Tsinghua University, Beijing, China 100084. xxie@tsinghua.edu.cn.

Z. Zeng is with Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. zexian.zeng@northwestern.edu.

J. Li is with Department of Industrial and Systems Engineering, University of Wisconsin, Madison, WI 53706. jli252@wisc.edu.

C.H. Swartz is with University of Kentucky Chandler Medical Center, Lexington, KY 40506, USA. chswar2@email.uky.edu.

P. DePriest is with Baptist Memorial Healthcare Corporation, Memphis, TN 38120, USA. paul.dePriest@bmhcc.org.

978-1-4673-8183-3/15/$31.00 ©2015 IEEE

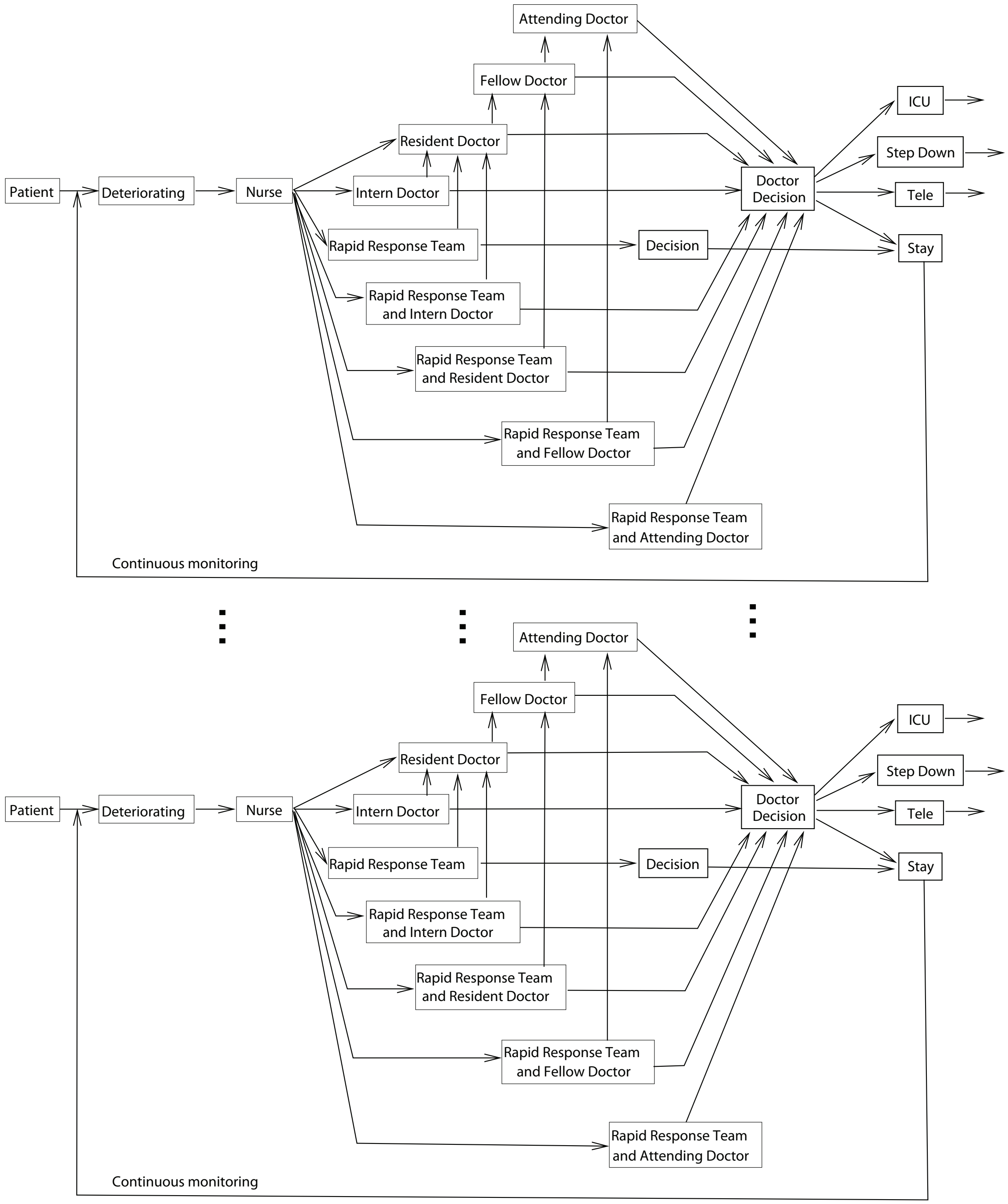

Fig. 1. Multi-patient rapid response process

patients' declining and responses from all providers, which cannot be estimated directly due to the complexity of the system. The problem in this paper is formulated as follows: Developing a method to evaluate the mean decision time.

III. Performance evaluation

When there is a single patient in the system, paper [8] presents the formula to evaluate the mean decision time T_d,

$$T_d = \sum_{i \in X_1 \cup X_2} p_i \tau_i, \tag{1}$$

where τ_i, $i \in X_1 \cup X_2$, is the average response time of each provider in X_1 or the two joint providers in X_2, and p_i is the probability that such response i, $i \in X_1 \cup X_2$, has been carried out. The calculation of p_i is presented in [8].

The single patient case assumes that the providers are always available. In the multiple patients case, a provider may not be available if there are more than one patients are declining, so that additional waiting is possible. To study such scenarios, we first consider an example of two patients, then extend to general cases.

A. A Two-Patient Example

Since when multiple patients are declining simultaneously, the limited resource may need to be shared (e.g., both patients need help from the RRT). Due to system complexity, a directly evaluation of decision time is all but impossible. Therefore, an iterative approach is pursued. Since there are two factors contributing to the waiting time: simultaneous patients declining and requesting for the same provider, a two-level iteration method is introduced. As shown in Figure 2, the Level-1 iteration investigates the possibility that the same resource is requested by more than one declining patients at the same time and quantifies the extra waiting time due to this. Thus, the mean decision time (including the extra waiting time) is obtained, denoted as T_{in}. Using this result, as well as the time the patient is in normal condition T_{normal}, the Level-2 iteration evaluates the probability that more than one patients are declining and the resulting waiting time by considering this probability.

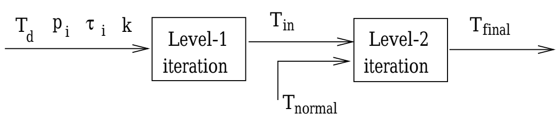

Fig. 2. Illustration of two-level iteration process

Specifically, in Level-1 iteration, using the single patient model, we obtain the mean decision time under complete availability of providers, T_d, and each provider's probability to be requested for help, p_i, $i \in X_1 \cup X_2$. Consider patient k, $k = 1, 2$, and resource r, $r \in X_1 \cup X_2$. Let $\tau_{k,r}^{(j)}$ denote the mean decision time which includes patient k's waiting time for resource r during the j-th iteration, $j = 1, 2, \ldots$. Introduce $p_{k,r}^{(j)}$ to denote the probability that patient k is occupying resource r when there is another request for the same resource during the j-th iteration. At the beginning of the iteration, let all $\tau_{k,r}^{(0)}$ equal to T_d and $p_{k,r}^{(0)}$ equal to 0. For patient 1, the waiting will happen if he/she requests help from intern while patient 2 is being attended by the intern. This probability can be expressed as $p_{2,int}^{(0)}$. In addition, the average time that the intern provides rapid response can be expressed by $p_{int}\tau_{int} + p_{rrt\&int}\tau_{rrt\&int}$. This is due to the intern's role in both single provider (*int*) and joint providers ($rrt\&int$) cases. Therefore, we update the mean decision time by including patient 1's waiting for intern, which is $\tau_{1,int}^{(1)}$, as follows:

$$\tau_{1,int}^{(1)} = T_d + p_{2,int}^{(0)}(p_{int}\tau_{int} + p_{rrt\&int}\tau_{rrt\&int}).$$

Using this $\tau_{1,int}^{(1)}$, we next update $p_{1,int}^{(1)}$, the probability that patient 1 is occupying the intern when there is another request for the resource. Again request for intern can happen in two cases: the single provider and joint providers. For the first case, $p_{int}\tau_{int}/\tau_{1,int}^{(1)}$ reflects the percentage of time the intern is occupied. Multiplying it by p_{int}, the probability the intern is working with another patient, we obtain the desired probability. Similarly, for the latter case, we obtain $p_{rrt\&int}^2\tau_{rrt\&int}/\tau_{1,int}^{(1)}$. Therefore, we have,

$$p_{1,int}^{(1)} = \frac{p_{int}^2\tau_{int} + p_{rrt\&int}^2\tau_{rrt\&int}}{\tau_{1,int}^{(1)}}.$$

Then, using $p_{1,int}^{(1)}$, the mean decision time contributed by patient 2's waiting for the intern, $\tau_{2,int}^{(1)}$, and the probability that patient 2 is occupying intern when the intern is requested by another patient, $p_{2,int}^{(1)}$, are updated.

$$\tau_{2,int}^{(1)} = T_d + p_{1,int}^{(1)}(p_{int}\tau_{int} + p_{rrt\&int}\tau_{rrt\&int}),$$
$$p_{2,int}^{(1)} = \frac{p_{int}^2\tau_{int} + p_{rrt\&int}^2\tau_{rrt\&int}}{\tau_{2,int}^{(1)}}.$$

This finishes the update related to the intern.

Next, we consider another resource, the resident, with similar updating process from patient 1 to patient 2. Specifically, for patient 1:

$$\tau_{1,res}^{(1)} = T_d + p_{2,res}^{(0)}(p_{res}\tau_{res} + p_{rrt\&res}\tau_{rrt\&res}),$$
$$p_{1,res}^{(1)} = \frac{p_{res}^2\tau_{res} + p_{rrt\&res}^2\tau_{rrt\&res}}{\tau_{1,res}^{(1)}}.$$

For patient 2:

$$\tau_{2,res}^{(1)} = T_d + p_{1,res}^{(1)}(p_{res}\tau_{res} + p_{rrt\&res}\tau_{rrt\&res}),$$
$$p_{2,res}^{(1)} = \frac{p_{res}^2\tau_{res} + p_{rrt\&res}^2\tau_{rrt\&res}}{\tau_{2,res}^{(1)}}.$$

Note that the current update of $p_{1,res}^{(1)}$ is used to obtain $\tau_{2,res}^{(1)}$. Then we continue to update all the rest of resources as follows to complete the first iteration. For RRT: from patients 1 to 2, we have

$$\begin{aligned}
\tau_{1,rrt}^{(1)} &= T_d + p_{2,rrt}^{(0)}(p_{rrt}\tau_{rrt} + p_{rrt\&int}\tau_{rrt\&int} \\
&\quad + p_{rrt\&res}\tau_{rrt\&res} + p_{rrt\&f}\tau_{rrt\&f} + p_{rrt\&a}\tau_{rrt\&a}), \\
p_{1,rrt}^{(1)} &= (p_{rrt}^2\tau_{rrt} + p_{rrt\&int}^2\tau_{rrt\&int}^2 + p_{rrt\&res}^2\tau_{rrt\&res} \\
&\quad + p_{rrt\&f}^2\tau_{rrt\&f} + p_{rrt\&a}^2\tau_{rrt\&a})/\tau_{1,rrt}^{(1)}, \\
\tau_{2,rrt}^{(1)} &= T_d + p_{1,rrt}^{(1)}(p_{rrt}\tau_{rrt} + p_{rrt\&int}\tau_{rrt\&int} \\
&\quad + p_{rrt\&res}\tau_{rrt\&res} + p_{rrt\&f}\tau_{rrt\&f} + p_{rrt\&a}\tau_{rrt\&a}), \\
p_{2,rrt}^{(1)} &= (p_{rrt}^2\tau_{rrt} + p_{rrt\&int}^2\tau_{rrt\&int}^2 + p_{rrt\&res}^2\tau_{rrt\&res} \\
&\quad + p_{rrt\&f}^2\tau_{rrt\&f} + p_{rrt\&a}^2\tau_{rrt\&a})/\tau_{2,rrt}^{(1)},
\end{aligned}$$

For fellow, from patients 1 to 2, we have

$$\tau_{1,f}^{(1)} = T_d + p_{2,f}^{(0)}(p_f\tau_f + p_{rrt\&f}\tau_{rrt\&f}),$$
$$p_{1,f}^{(1)} = \frac{p_f^2\tau_f + p_{rrt\&f}^2\tau_{rrt\&f}}{\tau_{1,f}^{(1)}},$$
$$\tau_{2,f}^{(1)} = T_d + p_{1,f}^{(1)}(p_f\tau_f + p_{rrt\&f}\tau_{rrt\&f}),$$
$$p_{2,f}^{(1)} = \frac{p_f^2\tau_f + p_{rrt\&f}^2\tau_{rrt\&f}}{\tau_{2,f}^{(1)}}.$$

Finally, for attending, addressing patients 1 to 2, we have

$$\tau_{1,a}^{(1)} = T_d + p_{2,a}^{(0)}(p_a\tau_a + p_{rrt\&a}\tau_{rrt\&a}),$$
$$p_{1,a}^{(1)} = \frac{p_a^2\tau_a + p_{rrt\&a}^2\tau_{rrt\&a}}{\tau_{1,a}^{(1)}},$$
$$\tau_{2,a}^{(1)} = T_d + p_{1,a}^{(1)}(p_a\tau_a + p_{rrt\&a}\tau_{rrt\&a}),$$
$$p_{2,a}^{(1)} = \frac{p_a^2\tau_a + p_{rrt\&a}^2\tau_{rrt\&a}}{\tau_{2,a}^{(1)}}.$$

After finishing the first iteration, the above updated variables are used for subsequent iterations and the process is repeated until it converges. More specifically, the following criteria is used for convergence check: Let $\delta = 10^{-5}$, if

$$|\tau_{i,r}^{(j+1)} - \tau_{i,r}^{(j)}| \leq \delta, \quad |p_{i,r}^{(j+1)} - p_{i,r}^{(j)}| \leq \delta,$$
$$i = 1, 2, \quad r \in X_1,$$

then we claim the procedure is convergent, and we denote

$$\lim_{j\to\infty} \tau_{i,r}^{(j)} = \tau_{i,r}, \quad \lim_{j\to\infty} p_{i,r}^{(j)} = p_{i,r}, \quad i = 1, 2.$$

Moreover, all $\tau_{i,r}, i = 1, 2$, should be identical and all $p_{i,r}, i = 1, 2$, will be the same, since we do not distinguish the patients. Therefore, upon convergence, we obtain the mean decision time (including waiting time) T_r and provider utilization probability P_r

$$\tau_{1,r} = \tau_{2,r} := \mathrm{T}_r, \qquad p_{1,r} = p_{2,r} := P_r.$$

Then the output of Level-1 iteration, T_{in}, can be obtained, which is the updated mean decision time by including the extra waiting time for all resources.

$$T_{in} = T_d + \Sigma_{r,r\in X_1} P_r \mathrm{T}_r. \tag{2}$$

Using the T_{in} from Level-1 iteration, we start the Level-2 iteration. Let $g_k^{(l)}$, $k = 1, 2$, $l = 1, 2, \ldots$, denote the time portion the patient is in declining status in iteration *j*, and $\mu_k^{(l)}$, $k = 1, 2$, $l = 1, 2, \ldots$, characterize the newly updated mean decision time in iteration *j* after considering the time proportion that patient *k* is in declining status. To start the iteration, assume all $g_k^{(0)}$ equal to 0 and $\mu_k^{(0)}$ to T_{in}. In the first iteration, for patient 1, the waiting will only occur if patient 1 is declining, while patient 2 is also declining, which can be expressed as $g_1^{(0)} g_2^{(0)}$. Therefore we first obtain the updated $\mu_1^{(1)}$ as follows.

$$\mu_1^{(1)} = T_{in}(1 + g_1^{(0)} g_2^{(0)}).$$

In addition, the time portion patient 1 is in declining status is updated as

$$g_1^{(1)} = \frac{\mu_1^{(1)}}{\mu_1^{(1)} + T_{normal}}.$$

Applying the updated $g_1^{(1)}$ and moving on to patient 2 with the same logic, we obtained the new $\mu_2^{(1)}$ and $g_2^{(1)}$ below:

$$\mu_2^{(1)} = T_{in}(1 + g_2^{(0)} g_1^{(1)}),$$
$$g_2^{(1)} = \frac{\mu_2^{(1)}}{\mu_2^{(1)} + T_{normal}}.$$

This finishes the first iteration. Using the results of $g_i^{(1)}$ and $\mu_i^{(1)}$, we continue this process by considering patients 1 and 2 again to update $g_i^{(l)}$ and $\mu_i^{(l)}$, and repeat the process until the procedure converges. The convergence criteria is met when the following conditions hold:

$$|\mu_i^{(j+1)} - \mu_i^{(l)}| \leq \delta, \quad |g_i^{(j+1)} - g_i^{(l)}| \leq \delta, \quad i = 1, 2,$$

again $\delta = 10^{-5}$. If the procedure converges, we obtain

$$\lim_{l\to\infty} \mu_i^{(l)} = \mu_i, \quad \lim_{l\to\infty} g_i^{(l)} = g_i, \quad i = 1, 2.$$

Same as in Level-1 iteration, all $\mu_i, i = 1, 2$, should be the same. Therefore, let T_{final} denote the final mean decision time, we have

$$\mu_1 = \mu_2 = T_{final}. \tag{3}$$

The above iterations address all the issues related to resource sharing, which contribute to the addition of extra waiting time. It can be proved that in both Level-1 and Level-2 iterations, the procedures are convergent.

B. Generalized Procedure

For the general case with n patients in the system, the iteration procedure can be formally described as Procedure 1. Due to the excessive length of the procedure, it is presented in the Appendix. As explained in Subsection III-A, such a procedure converges in the two-patient example. When the system has more than two patients, Level-2 iteration converges for any value of n.

Proposition 1: Level-2 iteration of Procedure 1 is convergent. The following limits exist:

$$\lim_{j\to\infty} \mu_i^{(l)} = \mu_i, \quad \lim_{j\to\infty} g_i^{(l)} = g_i, \quad i = 1, \ldots, n. \tag{4}$$

However, the convergence of Level-1 iteration can only be proved for the $n = 2$ case.

Proposition 2: When $n = 2$, Level-1 iteration of Procedure 1 is convergent. The following limits exist:

$$\lim_{j\to\infty} \tau_{i,r}^{(j)} = \tau_{i,r}, \quad \lim_{j\to\infty} p_{i,r}^{(j)} = p_{i,r}, \quad i = 1, 2, \quad r \in X_1. \tag{5}$$

To investigate the convergence in the general case, extensive numerical investigation using randomly generated parameters has been carried out and the results show that the procedure always converges with a unique solution. An illustration of convergence of $\tau_{i,res}$ in Level-1 iteration is shown in Figure 3. The convergence of other variables is similar. In addition, Figure 4 illustrates the convergence of g_i in Level-2 iteration.

As one can see, in both cases, the procedure converges within 3 iterations. In all the examples we tested, the convergence is always guaranteed, within 3-5 iterations.

IV. Accuracy and Applicability

The accuracy of Procedure 1 has been examined by comparing with the simulation results for dozens examples. In all experiments, the response time of each provider follows a uniform distribution between 20 and 40 minutes. All routing probabilities are uniformed selected from 0 to 1. The time

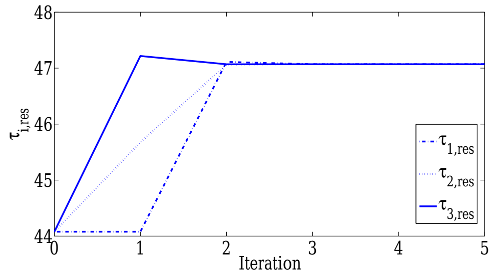

Fig. 3. Convergence of Level-1 iteration: resident doctor

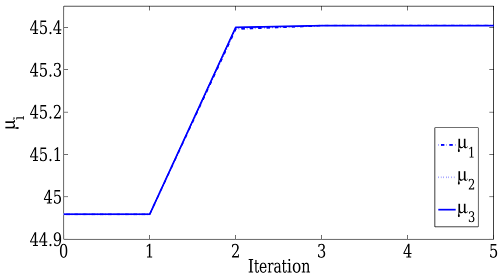

Fig. 4. Convergence of Level-2 iteration: μ

distribution of patient in normal status is assumed to be exponential. The simulation time and number of replications are selected to ensure that the confidence intervals are within 1% of the mean decision time. Let $T_{final}^{sim,i}$ and $T_{final}^{iter,i}$ denote the mean decision time obtained by simulation and by iteration method for experiment i. Then the relative error between $T_{final}^{sim,i}$ and $T_{final}^{iter,i}$ is defined as ϵ,

$$\epsilon_i = \frac{|T_{final}^{sim,i} - T_{final}^{iter,i}|}{T_{final}^{sim,i}} \cdot 100\%.$$

The mean of ϵ_i, denoted as $\bar{\epsilon}$, defines the average relative error. It is shown that $\bar{\epsilon}$ is less than 2%. The error may come from the heuristic updates of the probabilities in the iterations. Since the error is small, the iteration procedure can provide an accurate estimation of decision time.

In the above studies, exponential distribution is assumed for the time duration when a patient is in normal status. In practice, such an assumption may not hold. Thus, investigating the impact of non-exponential distribution of a patient's normal time is of importance. To do this, gamma and lognormal distributions are used due to the fact that they both have two parameters and can place the CV with freedom. Since in rapid response process, the more time elapses, the more likely the patient will be at the risk of deterioration, which lead to smaller coefficient of variation (CV). Therefore, we focus on the cases where CV is less than 1.

Specifically, five data points, $CV = 0.25, 0.5, 0.75$, as well as $CV = 0$ and $CV = 1$ are considered. Lognormal distribution is assumed first. We hypothesize that the variability has little effect on the mean decision time. A dozen instances are created and the largest possible relative error is recorded. Let $T_{i,j}$ denote the mean decision time obtained from simulation for experiment j for the i-th CV value, where $i = 1, 2, 3, 4, 5$ correspond to $CV = 0, 0.25, 0.5, 0.75, 1$, respectively. Then δ_j defines the difference between the largest and smallest decision time in experiment j.

$$\delta_j = \frac{\max_i T_{i,j} - \min_i T_{i,j}}{\min_i T_{i,j}} \cdot 100\%.$$

By calculating the average of δ_j, denoted as $\bar{\delta}$, the impact of non-exponential normal time can be studied. It is shown that $\bar{\delta}$ is less than 0.5%. Similar results are obtained by assuming gamma distribution. Thus, the hypothesis is justified. The proposed iterative method can provide an accurate estimate of mean decision time in multi-patient scenario.

V. Conclusions

In this paper, a rapid response system with multiple patients is studied. A recursive method is introduced to estimate the mean decision time for deteriorating patients. The method is based on two-level iterations to calculate the additional waiting time due to limited availability of care providers. It is shown that the iterative procedures are convergent, and the accuracy is high. Therefore, this approach can provide a quantitative tool to for performance evaluation.

The future work can be directed to studying the scenario with more complex routing possibilities, evaluating the variability in decision time, and developing a continuous improvement method through identification the most critical constraints (i.e., bottlenecks) in the system.

Appendix: Iteration Procedure

Procedure 1: (1) Level-1 iteration

Step 1.1: Initialization: Calculate p_i, $i \in X = X_1 \cup X_2$, and T_d from [8]. Let $j = 0$. Set $\tau_{k,i}^{(j)}$ and $p_{k,i}^{(j)}$ equal to 0.

Step 1.2: Update $\tau_{k,i}^{(j)}$ and $p_{k,i}^{(j)}$: For patient 1,

$$\begin{aligned}
\tau_{1,int}^{(j+1)} &= T_d + \Sigma_{i=2}^{n} p_{i,int}^{(j)} (p_{int}\tau_{int} + p_{rrt\&int}\tau_{rrt\&int}),\\
p_{1,int}^{(j+1)} &= (p_{int}^2\tau_{int} + p_{rrt\&int}^2\tau_{rrt\&int})/\tau_{1,int}^{(j+1)},\\
\tau_{1,res}^{(j+1)} &= T_d + \Sigma_{i=2}^{n} p_{i,res}^{(j)} (p_{res}\tau_{res} + p_{rrt\&res}\tau_{rrt\&res}),\\
p_{1,res}^{(j+1)} &= (p_{res}^2\tau_{res} + p_{rrt\&res}^2\tau_{rrt\&res})/\tau_{1,res}^{(j+1)},\\
\tau_{1,rrt}^{(j+1)} &= T_d + \Sigma_{i=2}^{n} p_{i,rrt}^{(j)} (p_{rrt}\tau_{rrt} + p_{rrt\&int}\tau_{rrt\&int}\\
&\quad + p_{rrt\&res}\tau_{rrt\&res} + p_{rrt\&f}\tau_{rrt\&f}\\
&\quad + p_{rrt\&a}\tau_{rrt\&a}),\\
p_{1,rrt}^{(j+1)} &= (p_{rrt}^2\tau_{rrt} + p_{rrt\&int}^2\tau_{rrt\&int}^2 + p_{rrt\&res}^2\tau_{rrt\&res}\\
&\quad + p_{rrt\&f}^2\tau_{rrt\&f} + p_{rrt\&a}^2\tau_{rrt\&a})/\tau_{1,rrt}^{(j+1)},\\
\tau_{1,f}^{(j+1)} &= T_d + \Sigma_{i=2}^{n} p_{i,f}^{(j)} (p_f\tau_f + p_{rrt\&f}\tau_{rrt\&f}),\\
p_{1,f}^{(j+1)} &= (p_f^2\tau_f + p_{rrt\&f}^2\tau_{rrt\&f})/\tau_{1,f}^{(j+1)},\\
\tau_{1,a}^{(j+1)} &= T_d + \Sigma_{i=2}^{n} p_{i,a}^{(j)} (p_a\tau_a + p_{rrt\&a}\tau_{rrt\&a}),\\
p_{1,a}^{(j+1)} &= (p_a^2\tau_a + p_{rrt\&a}^2\tau_{rrt\&a})/\tau_{1,a}^{(j+1)}.
\end{aligned}$$

For patient $k = 2, \ldots, n-1$,

$$\begin{aligned}
\tau_{k,int}^{(j+1)} &= T_d + (\Sigma_{i=1}^{k-1} p_{i,int}^{(j+1)} + \Sigma_{i=k+1}^{n} p_{i,int}^{(j)})\\
&\quad \cdot (p_{int}\tau_{int} + p_{rrt\&int}\tau_{rrt\&int}),\\
p_{k,int}^{(j+1)} &= (p_{int}^2\tau_{int} + p_{rrt\&int}^2\tau_{rrt\&int})/\tau_{k,int}^{(j+1)},
\end{aligned}$$

$$
\begin{aligned}
\tau_{k,res}^{(j+1)} &= T_d + (\Sigma_{i=1}^{k-1} p_{i,res}^{(j+1)} + \Sigma_{i=k+1}^{n} p_{i,res}^{(j)}) \\
&\quad \cdot (p_{res}\tau_{res} + p_{rrt\&res}\tau_{rrt\&res}), \\
p_{k,res}^{(j+1)} &= (p_{res}^2\tau_{res} + p_{rrt\&res}^2\tau_{rrt\&res})/\tau_{k,res}^{(j+1)}, \\
\tau_{k,rrt}^{(j+1)} &= T_d + (\Sigma_{i=1}^{k-1} p_{i,rrt}^{(j+1)} + \Sigma_{i=k+1}^{n} p_{i,rrt}^{(j)}) \\
&\quad \cdot (p_{rrt}\tau_{rrt} + p_{rrt\&int}\tau_{rrt\&int} + p_{rrt\&res}\tau_{rrt\&res} \\
&\quad + p_{rrt\&f}\tau_{rrt\&f} + p_{rrt\&a}\tau_{rrt\&a}), \\
p_{k,rrt}^{(j+1)} &= (p_{rrt}^2\tau_{rrt} + p_{rrt\&int}^2\tau_{rrt\&int}^2 + p_{rrt\&res}^2\tau_{rrt\&res} \\
&\quad + p_{rrt\&f}^2\tau_{rrt\&f} + p_{rrt\&a}^2\tau_{rrt\&a})/\tau_{k,rrt}^{(j+1)}, \\
\tau_{k,f}^{(j+1)} &= T_d + (\Sigma_{i=1}^{k-1} p_{i,f}^{(j+1)} + \Sigma_{i=k+1}^{n} p_{i,f}^{(j)}) \\
&\quad \cdot (p_f\tau_f + p_{rrt\&f}\tau_{rrt\&f}), \\
p_{k,f}^{(j+1)} &= (p_f^2\tau_f + p_{rrt\&f}^2\tau_{rrt\&f})/\tau_{k,f}^{(j+1)}, \\
\tau_{k,a}^{(j+1)} &= T_d + (\Sigma_{i=1}^{k-1} p_{i,a}^{(j+1)} + \Sigma_{i=k+1}^{n} p_{i,a}^{(j)}) \\
&\quad \cdot (p_a\tau_a + p_{rrt\&a}\tau_{rrt\&a}), \\
p_{k,a}^{(j+1)} &= (p_a^2\tau_a + p_{rrt\&a}^2\tau_{rrt\&a})/\tau_{k,a}^{(j+1)}.
\end{aligned}
$$

For patient $k = n$,

$$
\begin{aligned}
\tau_{k,int}^{(j+1)} &= T_d + \Sigma_{i=1}^{n-1} p_{i,int}^{(j+1)}(p_{int}\tau_{int} \\
&\quad + p_{rrt\&int}\tau_{rrt\&int}), \\
p_{k,int}^{(j+1)} &= \frac{p_{int}^2\tau_{int} + p_{rrt\&int}^2\tau_{rrt\&int}}{\tau_{k,r}^{(j+1)}}, \\
\tau_{k,res}^{(j+1)} &= T_d + \Sigma_{i=1}^{n-1} p_{i,res}^{(j+1)}(p_{res}\tau_{res} \\
&\quad + p_{rrt\&res}\tau_{rrt\&res}), \\
p_{k,res}^{(j+1)} &= (p_{res}^2\tau_{res} + p_{rrt\&res}^2\tau_{rrt\&res})/\tau_{k,res}^{(j+1)}, \\
\tau_{k,rrt}^{(j+1)} &= T_d + \Sigma_{i=1}^{n-1} p_{i,rrt}^{(j+1)}(p_{rrt}\tau_{rrt} \\
&\quad + p_{rrt\&int}\tau_{rrt\&int} + p_{rrt\&res}\tau_{rrt\&res} \\
&\quad + p_{rrt\&f}\tau_{rrt\&f} + p_{rrt\&a}\tau_{rrt\&a}), \\
p_{k,r}^{(j+1)} &= (p_{rrt}^2\tau_{rrt} + p_{rrt\&int}^2\tau_{rrt\&int}^2 \\
&\quad + p_{rrt\&res}^2\tau_{rrt\&res} + p_{rrt\&f}^2\tau_{rrt\&f} \\
&\quad + p_{rrt\&a}^2\tau_{rrt\&a})/\tau_{k,rrt}^{(j+1)}, \\
\tau_{k,f}^{(j+1)} &= T_d + \Sigma_{i=1}^{n-1} p_{i,f}^{(j+1)}(p_f\tau_f + p_{rrt\&f}\tau_{rrt\&f}), \\
p_{k,f}^{(j+1)} &= (p_f^2\tau_f + p_{rrt\&f}^2\tau_{rrt\&f})/\tau_{k,f}^{(j+1)}, \\
\tau_{k,a}^{(j+1)} &= T_d + \Sigma_{i=1}^{n-1} p_{i,a}^{(j+1)}(p_a\tau_a + p_{rrt\&a}\tau_{rrt\&a}), \\
p_{k,a}^{(j+1)} &= (p_a^2\tau_a + p_{rrt\&a}^2\tau_{rrt\&a})/\tau_{k,a}^{(j+1)}.
\end{aligned}
$$

Step 1.3: Iteration: Let $j = j + 1$. Go back to Step 1.2 until all the stopping criteria is met. For a given $\delta = 10^{-5}$, we terminate the Level-1 iteration until:

$$|\tau_{i,r}^{(j+1)} - \tau_{i,r}^{(j)}| \le \delta, \quad |p_{i,r}^{(j+1)} - p_{i,r}^{(j)}| \le \delta, \quad i = 1, 2, \ldots, n.$$

Step 1.4: Termination: When the above terminating conditions are met, let

$$
\begin{aligned}
&\tau_{i,r} = \mathrm{T}_r, \qquad p_{i,r} = P_r, \qquad i = 1, \ldots, n, \\
&T_{in} = T_d + \Sigma_{r,r\in X_1} P_r \mathrm{T}_r.
\end{aligned}
$$

(2) Level-2 iteration

Step 2.1: Initialization: Let $l = 0$. Set $g_1^{(l)} = 0$ and $\mu_1^{(l)} = T_{in}$.

Step 2.2: Update $g_k^{(l)}$ and $\mu_k^{(l)}$: For patient 1,

$$
\begin{aligned}
\mu_1^{(l+1)} &= T_{in}(1 + g_1^{(l)}\Sigma_{i=2}^{n} g_i^{(l)}), \\
g_1^{(l+1)} &= \frac{\mu_1^{(l+1)}}{\mu_1^{(l+1)} + T_{normal}}.
\end{aligned}
$$

For patient $k = 2, \ldots, n-1$,

$$
\begin{aligned}
\mu_k^{(l+1)} &= T_{in}(1 + g_k^{(l)}(\Sigma_{i=1}^{k-1} g_i^{(l+1)} + \Sigma_{i=k+1}^{n} g_i^{(l)})), \\
g_k^{(l+1)} &= \frac{\mu_k^{(l+1)}}{\mu_k^{(l+1)} + T_{normal}}.
\end{aligned}
$$

For patient $k = n$,

$$
\begin{aligned}
\mu_k^{(l+1)} &= T_{in}(1 + g_k^{(l)}\Sigma_{i=1}^{n-1} g_i^{(l+1)}), \\
g_k^{(l+1)} &= \frac{\mu_k^{(l+1)}}{\mu_k^{(l+1)} + T_{normal}}.
\end{aligned}
$$

Step 2.3: Iteration: Let $l = l + 1$. Go back to Step 2.2 until all the stopping criteria is met.

$$|\mu_i^{(l+1)} - \mu_i^{(l)}| \le \delta, \quad |g_i^{(l+1)} - g_i^{(l)}| \le \delta, \quad i = 1, 2, \ldots, n.$$

Step 2.4: Termination: When the stopping condition is met, we have

$$\mu_i^{(l+1)} = \mu_i, \quad \mu_i = T_{final}, \quad i = 1, \ldots, n.$$

References

[1] D.M. Berwick, D.R. Calkins, C.J. McCannon and A.D. Hackbarth, "The 100 000 Lives Campaign: Setting a Goal and a Deadline for Improving Health Care Quality," *The Journal of the American Medical Association*, vol. 295, pp. 324-327, 2006.

[2] L.T. Kohn, J.M. Corrigan and M.S. Donaldson, "To Err Is Human: Building a Safer Health System," Institute of Medicine, National Academy Press, Washington DC, 2000.

[3] D. Massey, L. M. Aitken and W. Chaboyer, "Literature Review: Do Rapid Response Systems Reduce the Incidence of Major Adverse Events in the Deteriorating Ward Patient?" *Journal of Clinical Nursing*, vol. 19, pp. 3260-3273, 2010.

[4] M.A. DeVita, R. Bellomo, K. Hillman, J. Kellum, A. Rotondi, D. Teres, A. Auerbach, W.-J. Chen, K. Duncan and G. Kenward, "Findings of the First Consensus Conference on Medical Emergency Teams," *Critical Care Medicine*, vol. 34, pp. 2463-2478, 2006.

[5] K. Hillman, J. Chen, M. Cretikos, R. Bellomo, D. Brown, G. Doig, S. Finfer and A. Flabouris, "Introduction of the Medical Emergency Team (MET) System: A Cluster-randomised Controlled Trial," *Lancet*, vol. 365, pp. 2091-2097, 2005.

[6] M.J. Dacey, E.R. Mirza, V. Wilcox, M. Doherty, J. Mello, A. Boyer, J. Gates, T. Brothers and R. Baute, "The Effect of a Rapid Response Team on Major Clinical Outcome Measures in a Community Hospital," *Critical Care Medicine*, vol. 35, pp. 2076-2082, 2007.

[7] P.S. Chan, R. Jain, B.K. Nallmothu, R.A. Berg and C. Sasson, "Rapid Response Teams: A Systematic Review and Meta-analysis," *Archives of Internal Medicine*, vol. 170, pp. 18-26, 2010.

[8] X. Xie, J. Li, C. H. Swartz, and P. Depriest, "Modeling and Analysis of Rapid Response Process to Improve Patient Safety," *IEEE Transactions on Automation Science and Engineering*, vol. 9, no. 2, pp. 215-225, 2012.

[9] X. Xie, J. Li, C.H. Swartz and P. DePriest, "Improving Response-Time Performance in Acute Care Delivery: A Systems Approach," *IEEE Transactions on Automation Science and Engineering*, vol. 11, pp. 1240-1249, 2014.

[10] X. Xie, *Hospital Rapid Response System: Modeling, Analysis and Improvement*, PhD Thesis, Department of Industrial and Systems Engineering, University of Wisconsin, Madison, 2014.