Online Appendix A (Proofs)

Proof of Theorem 1: To show the result, we use an *interchange* argument; we show that if classes $uc \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}$ and $sl \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}$ are such that $p'_{uc}\,\hat{\theta}'_{uc}\,\hat{\mu}'_{uc} \geq p'_{sl}\,\hat{\theta}'_{sl}\,\hat{\mu}'_{sl}$, then it is (weakly) better to serve class uc than class sl when in state $(\underline{x}, \underline{y})$ with $x_{uc}, x_{sl} > 0$. This will also prove that the optimal policy will not idle the physician when there are one or more patients available in the rooms, since idling can be thought of serving an additional class, class 0, with $\hat{\theta}'_0 = \hat{\mu}'_0 = p'_0 = 0$ (see, for instance, Buyukkoc et al. (1985)). To show that it is (weakly) better to serve class uc than class sl, we first consider the problem in an N-period discounted cost setting with four parallel (one for each class of patients) $\cdot/M/k$ systems (to guarantee bounded transition rates for the purpose of uniformization) in place of the $\cdot/M/\infty$ test stage, and show that the results hold for any number of periods to go $n \in 1, 2, ..., N$. (Notice that using four parallel $\cdot/M/k$ systems removes the need for considering the sequence and the type of patients within the common queue.) Using a convergence argument, as $n \to \infty$, it then follows that the result is true for an infinite-horizon (and hence, average cost) scenario with the four k-server test system. Next, taking limit as $k \to \infty$, it follows that the result is true even when transition rates are not bounded due to the existence of the $\cdot/M/\infty$ stage.

Now consider the finite horizon discounted cost version of (1). With β denoting the discount factor, the optimal discounted cost when there are $n+1$ (uniformized) periods to go is

$$\begin{aligned} V^k_{n+1}(\underline{x}, \underline{y}) = \frac{1}{\psi_k} \Big[\underline{\hat{\theta}'} (\underline{x} + \underline{y})^T + \beta \big[\sum_{ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}} [\lambda'_{ij} V^k_n(\underline{x} + \underline{e}_{ij}, \underline{y}) + (y_{ij} \wedge k)\, \eta\, V^k_n(\underline{x} + \underline{e}_{ij}, \underline{y} - \underline{e}_{ij})] \\ + \min_{a \in \mathcal{A}(\underline{x})} \big\{ \sum_{ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}} \mathbb{1}_{\{a=ij\}} \hat{\mu}'_{ij} \, [p'_{ij}\, V^k_n(\underline{x} - \underline{e}_{ij}, \underline{y}) + (1 - p'_{ij}) V^k_n(\underline{x} - \underline{e}_{ij}, \underline{y} + \underline{e}_{ij})] \\ + \big(\psi_k - \sum_{ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}} [\lambda'_{ij} + (y_{ij} \wedge k)\, \eta + \mathbb{1}_{\{a=ij\}} \hat{\mu}'_{ij}]\big)\, V^k_n(\underline{x}, \underline{y}) \big\} \big] \Big], \end{aligned} \tag{EC.1}$$

or equivalently (grouping the terms related to control in the minimization and self-loop)

$$\begin{aligned} V^k_{n+1}(\underline{x}, \underline{y}) = \frac{1}{\psi_k} \Big[\underline{\hat{\theta}'} (\underline{x} + \underline{y})^T + \beta \big[\sum_{ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}} [\lambda'_{ij} V^k_n(\underline{x} + \underline{e}_{ij}, \underline{y}) + (y_{ij} \wedge k)\, \eta\, V^k_n(\underline{x} + \underline{e}_{ij}, \underline{y} - \underline{e}_{ij})] \\ - \max_{a \in \mathcal{A}(\underline{x})} \big\{ \sum_{ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}} \mathbb{1}_{\{a=ij\}} \hat{\mu}'_{ij} \, [p'_{ij} \Delta^y_{ij}\, V^k_n(\underline{x} - \underline{e}_{ij}, \underline{y}) + \Delta^{x,y}_{ij} V^k_n(\underline{x} - \underline{e}_{ij}, \underline{y} + \underline{e}_{ij})] \\ + \big(\psi_k - \sum_{ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}} [\lambda'_{ij} + (y_{ij} \wedge k)\, \eta]\big)\, V^k_n(\underline{x}, \underline{y}) \big\} \big] \Big], \end{aligned} \tag{EC.2}$$

where $\Delta^y_{ij}\, V^k_n(\underline{x}, \underline{y}) = V^k_n(\underline{x}, \underline{y} + \underline{e}_{ij}) - V^k_n(\underline{x}, \underline{y})$ and $\Delta^{x,y}_{ij} V^k_n(\underline{x}, \underline{y}) = V^k_n(\underline{x} + \underline{e}_{ij}, \underline{y} - \underline{e}_{ij}) - V^k_n(\underline{x}, \underline{y})$. Now let π ($\hat{\pi}$) be the policy that prescribes serving patients of class uc (sl) for every state $(\underline{x}, \underline{y})$ with $x_{uc}, x_{sl} > 0$ and in every period n. From (EC.2), to show that π is (weakly) better than $\hat{\pi}$ in every period, we need to show that the following property holds for every n and every state $(\underline{x}, \underline{y})$ with $x_{uc}, x_{sl} > 0$:

$$\hat{\mu}'_{uc} \, [p'_{uc} \Delta^y_{uc}\, V^{k,\pi}_n(\underline{x} - \underline{e}_{uc}, \underline{y}) + \Delta^{x,y}_{uc} V^{k,\pi}_n(\underline{x} - \underline{e}_{uc}, \underline{y} + \underline{e}_{uc})]$$

$$\geq \hat{\mu}'_{sl}\,[p'_{sl}\Delta^{y}_{sl}\,V^{k,\hat{\pi}}_{n}(\underline{x}-\underline{e}_{sl},\underline{y})+\Delta^{x,y}_{sl}V^{k,\hat{\pi}}_{n}(\underline{x}-\underline{e}_{sl},\underline{y}+\underline{e}_{sl})]. \tag{EC.3}$$

To show property (EC.3), we use induction on n. First, for $n=0$, the property trivially holds since $V^{\pi}_0(\cdot,\cdot)=V^{\hat{\pi}}_0(\cdot,\cdot)=0$. Next, suppose the property holds for n. We show that it will then also hold for $n+1$. To do so, we need to consider different cases based on the state (i.e., partitions of the state space). First, consider the case where $x_{uc}, x_{sl}\geq 2$. Using action $a=uc$ (policy π) in both states $(\underline{x}-\underline{e}_{uc},\underline{y})$ and $(\underline{x},\underline{y})$ to compute $V^{k,\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})$ and $V^{k,\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y})$ using (EC.2), and subtracting the results we have $\Delta^{y}_{uc}V^{k,\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y})=$

$$\begin{aligned}
\frac{1}{\psi_k}\Big[\hat{\theta}'_{uc}+\beta\big[\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}&[\lambda'_{ij}\Delta^{y}_{uc}V^{k,\pi}_{n}(\underline{x},\underline{y})+(y_{ij}\wedge k)\,\eta\,\Delta^{y}_{uc}V^{k,\pi}_{n}(\underline{x}+\underline{e}_{ij}-\underline{e}_{uc},\underline{y}-\underline{e}_{ij})]\\
&\qquad+\mathbb{1}_{\{y_{uc}<k\}}\,\eta\,V^{k,\pi}_{n}(\underline{x},\underline{y}-\underline{e}_{uc})\\
&-\hat{\mu}'_{uc}\,[p'_{uc}\Delta^{y}_{uc}\Delta^{y}_{uc}\,V^{k,\pi}_{n}(\underline{x}-2\underline{e}_{uc},\underline{y})+\Delta^{y}_{uc}\Delta^{x,y}_{uc}V^{k,\pi}_{n}(\underline{x}-2\underline{e}_{uc},\underline{y}+\underline{e}_{ij})]\\
&+\big(\psi_k-\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[\lambda'_{ij}+(y_{ij}\wedge k)\,\eta]\big)\,\Delta^{y}_{uc}V^{k,\pi}_{n}(\underline{x}-\underline{e}_{uc},\underline{y})-\mathbb{1}_{\{y_{uc}<k\}}\,\eta\,V^{k,\pi}_{n}(\underline{x}-\underline{e}_{uc},\underline{y})\big]\Big].
\end{aligned} \tag{EC.4}$$

Similarly, we can derive $\Delta^{x,y}_{uc}V^{k,\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})$ using (EC.2) and action $a=uc$ (policy π) in both states $(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})$ and $(\underline{x},\underline{y})$ and subtracting the results. Doing so we have $\Delta^{x,y}_{uc}V^{k,\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})=$

$$\begin{aligned}
\frac{1}{\psi_k}\Big[\beta\big[\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}&[\lambda'_{ij}\Delta^{x,y}_{uc}V^{k,\pi}_{n}(\underline{x},\underline{y}+\underline{e}_{uc})+(y^{+}_{ij}\wedge k)\,\eta\,\Delta^{x,y}_{uc}V^{k,\pi}_{n}(\underline{x}+\underline{e}_{ij}-\underline{e}_{uc},\underline{y}-\underline{e}_{ij}+\underline{e}_{uc})]\\
&\qquad-\mathbb{1}_{\{y_{uc}<k\}}\,\eta\,V^{k,\pi}_{n}(\underline{x},\underline{y})\\
&-\hat{\mu}'_{uc}\,[p'_{uc}\Delta^{x,y}_{uc}\Delta^{y}_{uc}\,V^{k,\pi}_{n}(\underline{x}-2\underline{e}_{uc},\underline{y}+\underline{e}_{uc})+\Delta^{x,y}_{uc}\Delta^{x,y}_{uc}V^{k,\pi}_{n}(\underline{x}-2\underline{e}_{uc},\underline{y}+\underline{e}_{ij}+\underline{e}_{uc})]\\
&+\big(\psi_k-\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[\lambda'_{ij}+(y^{+}_{ij}\wedge k)\,\eta]\big)\,\Delta^{x,y}_{uc}V^{k}_{n}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})+\mathbb{1}_{\{y_{uc}<k\}}\,\eta\,V^{k,\pi}_{n}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})\big]\Big],
\end{aligned} \tag{EC.5}$$

where $y^{+}_{ij}=y_{ij}$ for all $ij\neq uc\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}$, and $y^{+}_{uc}=y_{uc}+1$. In a similar way and by using action $a=sl$ (policy $\hat{\pi}$) in (EC.2), quantities $\Delta^{y}_{sl}\,V^{k,\hat{\pi}}_{n+1}(\underline{x}-\underline{e}_{sl},\underline{y})$ and $\Delta^{x,y}_{sl}V^{k,\hat{\pi}}_{n+1}(\underline{x}-\underline{e}_{sl},\underline{y}+\underline{e}_{sl})$ can be computed. Next, to check property (EC.3) for $n+1$, multiply (EC.4) by $p'_{uc}\,\hat{\mu}'_{uc}$, and (EC.5) by $\hat{\mu}'_{uc}$ and add up the results. Similarly, multiply $\Delta^{y}_{sl}\,V^{k,\hat{\pi}}_{n+1}(\underline{x}-\underline{e}_{sl},\underline{y})$ and $\Delta^{x,y}_{sl}V^{k,\hat{\pi}}_{n+1}(\underline{x}-\underline{e}_{sl},\underline{y}+\underline{e}_{sl})$ by $p'_{sl}\,\hat{\mu}'_{sl}$ and $\hat{\mu}'_{sl}$, respectively, and add up the results. Next, using the induction hypothesis and that $p'_{uc}\,\hat{\theta}'_{uc}\,\hat{\mu}'_{uc}\geq p'_{sl}\,\hat{\theta}'_{sl}\,\hat{\mu}'_{sl}$, after algebraic simplification it follows that

$$\begin{aligned}
&\hat{\mu}'_{uc}\,[p'_{uc}\Delta^{y}_{uc}\,V^{k,\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y})+\Delta^{x,y}_{uc}V^{k,\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})]\\
&\qquad-\hat{\mu}'_{sl}\,[p'_{sl}\Delta^{y}_{sl}\,V^{k,\hat{\pi}}_{n+1}(\underline{x}-\underline{e}_{sl},\underline{y})+\Delta^{x,y}_{sl}V^{k,\hat{\pi}}_{n+1}(\underline{x}-\underline{e}_{sl},\underline{y}+\underline{e}_{sl})]\geq 0,
\end{aligned} \tag{EC.6}$$

which establishes property (EC.3) for $n+1$ for the case where $x_{uc},x_{sl}\geq 2$. In a similar way, this property can be established for other cases (i.e., the remaining partition of the state space). Hence, the non-idling strict priority rule is optimal for all n. Next, taking the limit as $n\to\infty$ it follows that the finite horizon problem converges to the infinite horizon one both in policy and cost (see Sennott (1999) Proposition 4.3.1). Furthermore, the convergence of the policy of the infinite-horizon discounted cost problem to that of average cost can easily be established (see Sennott (1999) Corollary 7.5.10). Therefore, the underlying non-idling strict priority policy is optimal under the average cost setting indexed by k (i.e., with $\cdot/M/k$'s in place of the $\cdot/M/\infty$) for any finite k. Since the result is true for any k, a convergence argument can be used to show that the result holds for the original problem with $k=\infty$. Notice that the existence of an optimal stationary policy for the original CTMS (i.e., when $k=\infty$) follows from the results of Guo and Liu (2001). □

Proof of Proposition 1: The proof of part (i) follows directly from comparing (5) and (6). To show part (ii), notice that, using the result of part (i) for a special case where there is no misclassification error, prioritizing U (N) patients is optimal if, and only if, $\theta_U\,\mu_U \geq (\leq)\theta_N\,\mu_N$. Next, observe that $\theta'_U\,\mu'_U - \theta'_N\,\mu'_N \;= \big[\lambda_N\lambda_U\mu_N\mu_U(\theta_U\mu_U-\theta_N\mu_N)(1-\gamma_N-\gamma_U)\big]/\big[\big(\lambda_N\mu_U\gamma_N+\lambda_U\mu_N(1-\gamma_U)\big)\big(\lambda_N\mu_U(1-\gamma_N)+\lambda_U\mu_N\gamma_U\big)\big]$. Combining these two results completes the proof of part (ii), as the sign of the numerator changes when the sum of errors exceeds 1. □

LEMMA EC.1 **(Perfect Classification - Prioritization)**. *In the simplified single-stage ED model under perfect urgency and complexity based classification:*
(i) The best priority rule is to prioritize patients in decreasing order of $\theta\mu$ values. Hence, if $\theta_{UC}\mu_{UC} \geq \theta_{NS}\mu_{NS}$, then the best priority rule is to follow the ordering: US, UC, NS, NC. Otherwise, the ED should follow the priority ordering: US, NS, UC, NC.
(ii) $R_*^{\mathcal{U}\cup\mathcal{C}} \leq R_*^{\mathcal{U}}$. That is, the risk of adverse events under the optimal priority rule using both complexity and urgency information is (weakly) smaller than that under the optimal apriority rule using only urgency information.
(iii) The best priority rule of part (i) is optimal even among the larger class of all non-anticipative policies (state or history dependent, idling or non-idling, etc.).

Proof of Lemma EC.1: Notice that, using (2), we can compute the average waiting time of each class of patients under any (static) priority rule. Furthermore, under priority rule π, we have

$$R_\pi^{\mathcal{U}\cup\mathcal{C}} = \sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{C}} \theta_{ij}\lambda_{ij}W_{ij}^\pi, \tag{EC.7}$$

where W_{ij}^π is the average waiting of class ij under priority rule π. The proof of part (i) then follows from Cox and Smith (1961) (see pages 83-84), where an interchange argument is used (when the number of customer classes is at least 3) to show that the best rule (among the priority policies) to minimize the holding cost in a non-preemptive M/G/1 is to follow the $c\mu$ rule. Replacing holding cost values (c) with adverse event rates (θ), and noticing that the patient class US (NC) has the highest (lowest) $\theta\mu$ value complete the proof of part (i). Next, using the result of part (i) together with (2) and (EC.7), when $\theta_{UC}\mu_{UC} \geq \theta_{NS}\mu_{NS}$, we have:

$$R_*^{\mathcal{U}\cup\mathcal{C}} = \lambda\mathbb{E}(s^2)\Big[\frac{\lambda_{US}\theta_{US}}{2(1-\rho_{US})} + \frac{\lambda_{UC}\theta_{UC}}{2(1-\rho_{US})(1-\rho_{US}-\rho_{UC})} + \frac{\lambda_{NS}\theta_{NS}}{2(1-\rho_{US}-\rho_{UC})(1-\rho_{US}-\rho_{UC}-\rho_{NS})} + \frac{\lambda_{NC}\theta_{NC}}{2(1-\rho_{US}-\rho_{UC}-\rho_{NS})(1-\rho_{US}-\rho_{UC}-\rho_{NS}-\rho_{NC})}\Big] \quad \text{(EC.8)}$$
$$\le \min\{R_U^{\mathcal{U}}, R_N^{\mathcal{U}}\} = R_*^{\mathcal{U}},$$

where the inequality follows from (3) and (4) together with the result of part (i) of Proposition 1 (for the special case where there is no misclassification error).

When $\theta_{UC}\mu_{UC} < \theta_{NS}\mu_{NS}$, we have:

$$R_*^{\mathcal{U}\cup\mathcal{C}} = \lambda\mathbb{E}(s^2)\Big[\frac{\lambda_{US}\theta_{US}}{2(1-\rho_{US})} + \frac{\lambda_{NS}\theta_{NS}}{2(1-\rho_{US})(1-\rho_{US}-\rho_{NS})} + \frac{\lambda_{UC}\theta_{UC}}{2(1-\rho_{US}-\rho_{NS})(1-\rho_{US}-\rho_{NS}-\rho_{UC})} + \frac{\lambda_{NC}\theta_{NC}}{2(1-\rho_{US}-\rho_{UC}-\rho_{NS})(1-\rho_{US}-\rho_{UC}-\rho_{NS}-\rho_{NC})}\Big], \quad \text{(EC.9)}$$

and similar to the previous case, it can be easily seen that $R_*^{\mathcal{U}\cup\mathcal{C}} \le R_*^{\mathcal{U}}$. The proof of part (iii) follows from Kakalik and Little (1971) (after replacing holding cost with adverse event rates) who (for the average holding cost objective) showed that the $c\mu$ policy of Cox and Smith (1961) remains optimal even when inserting idleness is allowed and/or when the priority rule is dynamic (i.e., state-dependent). □

Proof of Proposition 2: The proof of part (i) follows directly from the proof of part (i) of Lemma EC.1, since all rates are replaced with their error impacted counterparts. That is, the same interchange method of Cox and Smith (1961) (see pages 83-84) after replacing all rates with their error impacted counterparts proves that the best priority rule is to give priority based on a decreasing order of $\theta\mu$ values. The proof of part (ii) follows from the proof of Lemma EC.1 (found earlier in this appendix) part (ii) after replacing parameters with their error impacted counter parts. The proof of part (iii) follows from the result of Kakalik and Little (1971), after replacing holding cost with the error impacted intensity of adverse events, and all the other rates with their error impacted counterparts. □

LEMMA EC.2 **(Perfect Classification - Attractiveness)**. *In the simplified single-stage ED model, perfect complexity-augmented triage yields a larger improvement over perfect urgency-based triage when (i) ED utilization is higher, (ii) heterogeneity in the average service time of simple vs. complex patients is larger, and/or (iii) the fraction of simple and complex patients are closer to equal.*

Proof of Lemma EC.2: To show the result, first consider the case where under the $\mathcal{U}\cup\mathcal{C}$ classification it is optimal to follow the priority order US, UC, NS, NC, and under the $\mathcal{U}$ classification, it is optimal to follow the priority order U, N (i.e., prioritizing urgent patients first). Let $f = R_*^{\mathcal{U}\cup\mathcal{C}} - R_*^{\mathcal{U}}$, and notice that with $\mu_{iC} = \mu_C$ and $\mu_{iS} = \mu_S$ ($\forall i \in \mathcal{U}$), and $\theta_{Uj} = \theta_U$ and $\theta_{Nj} = \theta_N$ ($\forall j \in \mathcal{C}$) (i.e., when complexity is based only on set $\mathcal{C}$ and urgency is based only on set $\mathcal{U}$), from (EC.8) and (3) we have:

$$f = -[\frac{\theta_U \lambda_{US} \lambda_{UC}(1/\mu_C - 1/\mu_S)}{2(1-\rho_U)} + \frac{\theta_N \lambda_{NC} \lambda_{NS}(1/\mu_C - 1/\mu_S)}{2(1-\rho_U)(1-\rho)}]. \tag{EC.10}$$

Then, a careful treatment of utilization (realizing that $\rho_U = \lambda_U/\mu_U$ and $\rho = \rho_U + \rho_N$) shows that f is non-increasing in utilization, ρ. To prove part (ii), it then can be seen that f is non-increasing in $1/\mu_C - 1/\mu_S$ (keeping utilization and other factors the same). To see part (iii), let $\alpha \in [0,1]$ denote the fraction of patients that are complex, and $(1-\alpha)$ denote the fraction of patients that are simple, so $\lambda_{US} = (1-\alpha)\lambda_U$, $\lambda_{UC} = \alpha\lambda_U$, $\lambda_{NC} = \alpha\lambda_N$, and $\lambda_{NS} = (1-\alpha)\lambda_N$. Replacing these in (EC.10), it follows that f, as a function of α, can be written as $f = -[\alpha(1-\alpha)]k$, for some constant $k \geq 0$. Thus, $\alpha = 0.5$ yields the maximum benefit. The proof for other cases (i.e., when other priority rules are optimal) follows a similar argument after computing f using either (EC.8) or (EC.9), and either (3) or (4), depending on the optimal priority rule under $\boldsymbol{\mathcal{U}} \cup \boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{U}}$ classifications, respectively. □

Proof of Proposition 3: The proof of parts (i) - (iii) follows mainly from the proof of Lemma EC.2. First, consider the case where under the $\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'$ (i.e., imperfect urgency and complexity) classification it is optimal to follow the priority order US, UC, NS, NC, and under the $\boldsymbol{\mathcal{U}}'$ (i.e., imperfect urgency) classification, it is optimal to follow the priority order U, N (i.e., prioritizing urgent patients over non-urgent patients). With $f = R_*^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'} - R_*^{\boldsymbol{\mathcal{U}}'}$, and after replacing rates with their error impacted counterparts in (EC.10) we have:

$$f = -[\frac{\theta'_U \lambda'_{US} \lambda'_{UC}(1/\mu'_C - 1/\mu'_S)}{2(1-\rho'_U)} + \frac{\theta'_N \lambda'_{NC} \lambda'_{NS}(1/\mu'_C - 1/\mu'_S)}{2(1-\rho'_U)(1-\rho')}]. \tag{EC.11}$$

Next, notice that $\rho' = \rho$ (i.e., the total utilizations with and without misclassifications are the same). Hence, similar to the proof of part (i) of Lemma EC.2, it can be seen that f is non-increasing in ρ. Moreover, it can be seen that f is non-increasing in $1/\mu'_C - 1/\mu'_S$. Next, notice that $(\underline{1}/\underline{\mu}')^T = (A\,(\underline{\lambda}/\underline{\mu})^T)/\underline{\lambda}'$, where A is defined in (7). Thus, similar to the proof of part (ii) of Lemma EC.2, it can be seen that f is non-increasing in $1/\mu_C - 1/\mu_S$, which proves part (ii). Furthermore, similarly to the proof of part (iii) of Lemma EC.2, let $\lambda_{US} = (1-\alpha)\lambda_U$, $\lambda_{UC} = \alpha\lambda_U$, $\lambda_{NC} = \alpha\lambda_N$, and $\lambda_{NS} = (1-\alpha)\lambda_N$. It can be seen that f as a function of α is minimized at $\alpha = 0.5$, which proves part (iii). It can also be seen that f is non-decreasing in complexity misclassification error rates, γ_S and γ_C, which proves part (iv). The proof for other cases (i.e., when other priority rules are optimal) follows a similar line of argument after computing f. □

Proof of Proposition 4: Note that since the total population of patients and workloads served in both designed is constant, having both streaming designs under consideration as perfectly balanced entails $\rho'_{ij} = \rho/4$ and $\lambda'_{ij} = \lambda/4$ (for all $ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}$). First consider the design where patients are streamed based on the complexity-based information (and prioritized based on the urgency-information). Label this design as D1. Using similar results to (5) and (6), and replacing λ with $\lambda/2$ (since the volume of patients in each stream are the same), for the stream of S patients under D1, ROAE denoted by R_S^{D1} is: $R_S^{D1} = \lambda \mathbb{E}(\tilde{s}^2)/4\Big[\theta'_{US}\lambda'_{US}/(1-\rho/4) + \theta'_{NS}\lambda'_{NS}/[(1-\rho/4)(1-\rho/2)]\Big]$. Similarly, for the stream of C patients under D1, ROAE denoted by R_C^{D1} is: $R_C^{D1} = \lambda \mathbb{E}(\tilde{s}^2)/4\Big[\theta'_{UC}\lambda'_{UC}/(1-\rho/4) + \theta'_{NC}\lambda'_{NC}/[(1-\rho/4)(1-\rho/2)]\Big]$. Furthermore, the total ROAE under D1 is $R^{D1} = R_S^{D1} + R_C^{D1}$. Next, consider the design where patients are streamed based on the urgency-based information (and prioritized based on complexity-based information). Label this design as D2, and notice that, for the stream of U patients under D2, ROAE denoted by R_U^{D2} is: $R_U^{D2} = \lambda \mathbb{E}(\tilde{s}^2)/4\Big[\theta'_{US}\lambda'_{US}/(1-\rho/4) + \theta'_{UC}\lambda'_{UC}/[(1-\rho/4)(1-\rho/2)]\Big]$. Similarly, for the stream of N patients under D2, ROAE denoted by R_N^{D2} is: $R_N^{D2} = \lambda \mathbb{E}(\tilde{s}^2)/4\Big[\theta'_{NS}\lambda'_{NS}/(1-\rho/4) + \theta'_{NC}\lambda'_{NC}/[(1-\rho/4)(1-\rho/2)]\Big]$. Also, the total ROAE under D2 is $R^{D2} = R_U^{D2} + R_N^{D2}$. Hence, after simplifications, we have:

$$R^{D1} - R^{D2} = \frac{\lambda\rho\mathbb{E}(\tilde{s}^2)}{8(1-\rho/4)(1-\rho/2)}[\theta'_{NS}\lambda'_{NS} - \theta'_{UC}\lambda'_{UC}],$$

which completes the proof after noticing that $\lambda'_{NS} = \lambda'_{UC}$ (as the streaming designs are assumed to be perfectly balanced), and (b) $\theta'_{NS} \le \theta'_{UC}$ (as the optimal priority rules in each stream are assumed to be based on the Proposition 2(i) and the streaming designs are assumed to be perfectly balanced). □

Online Appendix B (Comparison of Partially Balanced Streaming Designs)

In Section 5.3, we compared the performance of two streaming patient flow designs: streaming patients based on complexity information and prioritizing them based on urgency (complexity streaming), and streaming patients based on urgency and prioritizing them based on complexity (urgency streaming). We analytically showed that complexity streaming is preferred to urgency streaming when the streaming is done in a *perfectly balanced* manner. Here we consider the case where the streaming is only *partially balanced*; that is, we still assume that the utilizations in the two streams of each design can be balanced through appropriate capacity allocation but that volumes of patients sent to each stream, mean service times, service time variances will not be the same. We note that our collected data as well as some studies from the literature (e.g., Vance and Spirvulis (2005)) indicate that complexity can be defined in such a way that the volume of complex and simple patients are roughly equal. However, unlike complexity, in some ED's the percentage of urgent and non-urgent patients might not be roughly equal. Hence, we perform a sensitivity analysis on this factor. Because both the total population of patients and workloads served are constant in both designs, balancing the utilizations requires $\rho'_{ij} = \rho/4$ (for all $ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}$). Next, while we keep the volume of patients sent to both streams in complexity streaming equal (i.e., $\lambda'_{US} + \lambda'_{NS} = \lambda'_{UC} + \lambda'_{NC}$), we assume that the volume of patients sent to the non-urgent stream is $(1+\delta)$ times that of the urgent stream in urgency streaming and examine cases with $\delta \in \{-0.5, -0.25, 0, 0.25, 0.5\}$. After capacity allocation to make utilization equal in the two streams,

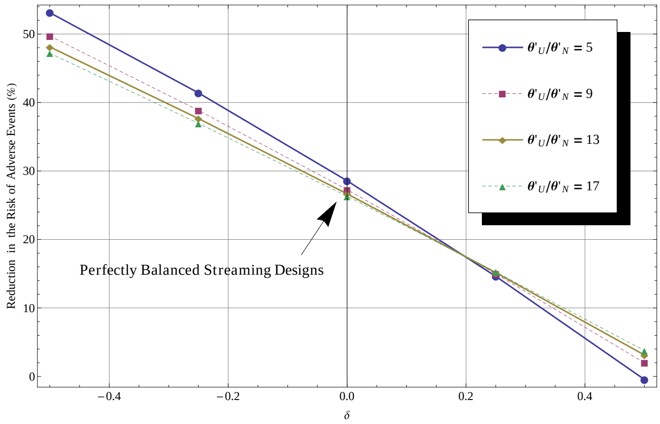

Figure EC.1 **Comparison of complexity streaming with urgency streaming.** $\delta \in [-0.5, 0.5]$ **represents deviation from perfectly balanced streaming systems (**$\theta'_{NC} = \theta'_{NS} = \theta'_N = 1, \theta'_{UC} = \theta'_{US} = \theta'_U$**).**

the effective (error-impacted) service rates are $\mu'_{US} = \mu'_{UC} = 1/2$ and $\mu'_{US} = \mu'_{UC} = (1+\delta) \times 1/2$. We further assume that the effect of capacity allocation on the service time variances is roughly equal to square of mean service times. Figure EC.1 depicts the percentage benefit of complexity streaming over urgency streaming for a numerical example with a total ED utilization of 80% where (error-impacted) arrivals are $\lambda'_{US} = \lambda'_{UC} = 1/10$ and $\lambda'_{NS} = \lambda'_{NC} = (1+\delta) \times 1/10$. As can be seen from Figure EC.1, complexity streaming is preferred to urgency streaming for these levels of δ over a wide range of values for the ratio of adverse event rates, θ'_U/θ'_N. Our other numerical examples and sensitivity analysis also show similar behavior, although we do not present them here for brevity.

Online Appendix C (General Occurrence of Adverse Events)

In the main body of the paper, we modeled the occurrence of adverse events with Poisson processes. We now relax the Poisson assumption and test the robustness of the Phase 1 priority rules proposed in the paper.

To this end, we allow the adverse events to occur based on any stationary point process (but not necessarily Poisson). For class $ij \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}$, let S'_{ij} and $\{\Theta'_{ij}(t), t \geq 0\}$ denote the error-impacted random service time and the error-impacted process that defines the number of adverse events until time t, respectively. Also, recall that $R_\pi^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t)$ denotes the total number of adverse events until time t under policy π, and consider the following notions of stochastic ordering for two random variables X and Y. If $Pr\{X > u\} \geq Pr\{Y > u\}$ for all $u \in \mathbb{R}$, then X is said to be *stochastically* greater than Y ($Y \leq_{st} X$). Similarly, we say that the counting process $\{N_1(t), t \geq 0\}$ is *stochastically* greater than the counting process $\{N_2(t), t \geq 0\}$ ($\{N_2(t), t \geq 0\} \leq_{st} \{N_1(t), t \geq 0\}$) if for all $t \geq 0$ $N_2(t) \leq_{st} N_1(t)$. If f and g represent the densities or probability mass functions of X and Y, respectively, and $f(\xi)/g(\xi)$ is increasing in ξ over the union of the supports of X and Y, then X is said to be greater than Y in the *likelihood ratio ordering* ($Y \leq_{lr} X$)(see, e.g., Shaked and Shanthikumar (2007) for more details). Finally, we let Ξ denote the class of all static priority rules, and use a sample path argument to show the following.

PROPOSITION EC.1. *If classes $ij, i'j' \in \boldsymbol{\mathcal{U}} \times \boldsymbol{\mathcal{C}}$ are such that $\{\Theta'_{i'j'}(t), t \geq 0\} \leq_{st} \{\Theta'_{ij}(t), t \geq 0\}$ and $S'_{ij} \leq_{lr} S'_{i'j'}$, then a policy that prioritizes patients of class $i'j'$ over those of class ij (whenever patients of both classes are available) cannot be uniquely optimal in the sense of stochastically minimizing $R_\pi^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t)$ at every $t \geq 0$ over all $\pi \in \Xi$.*

Proof of Proposition EC.1: We use a sample path argument. To this end, we first state the following result, a more general version of which can be found in Righter 1994 (Lemma 13.D.1) among others.

LEMMA EC.3. *Consider random variables X and Y, and let $\min\{X,Y\} = m$ and $\max\{X,Y\} = M$. If $X \leq_{lr} Y$, then $Pr\{X = m|m,M\} = Pr\{Y = M|m,M\} \geq Pr\{Y = m|m,M\} = Pr\{X = M|m,M\}$.*

Suppose policy $\pi \in \Xi$ takes a patient from class $i'j'$ at time t_0 while a patient of class ij is waiting in the waiting area. We construct a new policy, $\hat{\pi} \in \Xi$, which takes a patient of class ij at time t_0, and has a lower number of adverse events: $R_{\hat{\pi}}^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t) \leq R_\pi^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t)$ for all $t \geq 0$ along any sample path. To this end, using the above lemma, we employ a *cross coupling* (see, e.g., Section 13.D of Righter 1994) argument to prove the result as follows. Let S'^{γ}_l be the "error-impacted" service time of patients of class $l \in \{ij, i'j'\}$ (i.e., the service time for those classified as l) under policy $\gamma \in \{\pi, \hat{\pi}\}$. Using Lemma EC.3 (and since $S'_{ij} \leq_{lr} S'_{i'j'}$), we can cross couple $(S'^{\pi}_{ij}, S'^{\pi}_{i'j'})$ with $(S'^{\hat{\pi}}_{ij}, S'^{\hat{\pi}}_{i'j'})$ so that $m \triangleq \min\{S'^{\pi}_{ij}, S'^{\pi}_{i'j'}\} = \min\{S'^{\hat{\pi}}_{ij}, S'^{\hat{\pi}}_{i'j'}\}$, $M \triangleq \max\{S'^{\pi}_{ij}, S'^{\pi}_{i'j'}\} = \max\{S'^{\hat{\pi}}_{ij}, S'^{\hat{\pi}}_{i'j'}\}$, and either $S'^{\pi}_{i'j'} = S'^{\hat{\pi}}_{ij} \triangleq \alpha \in \{m, M\}$ and $S'^{\pi}_{ij} = S'^{\hat{\pi}}_{i'j'} \triangleq \beta \in \{m, M\} \backslash \{\alpha\}$ (Case 1) or $S'^{\pi}_{ij} = S'^{\hat{\pi}}_{ij} = m$ and $S'^{\pi}_{i'j'} = S'^{\hat{\pi}}_{i'j'} = M$ (Case 2). For all other random variables, we use the usual "direct coupling" mechanism, i.e., we use the same realizations under both policies. For the occurrence of adverse events, since $\{\Theta'_{ij}(t), t \geq 0\}$ and $\{\Theta'_{i'j'}(t), t \geq 0\}$ are based on stationary point processes and $\{\Theta'_{ij}(t), t \geq 0\} \geq_{st} \{\Theta'_{i'j'}(t), t \geq 0\}$, we note that along any sample path we have $\Theta'_{ij}(t_1 + \xi) - \Theta'_{ij}(t_1) = \Theta'_{ij}(\xi) \geq \Theta'_{i'j'}(\xi) = \Theta'_{i'j'}(t_1 + \xi) - \Theta'_{i'j'}(t_1)$ for any fixed times $t_1, \xi \geq 0$ (note that stochastic dominance results in sample path dominance). Let $\tilde{t} > t_0$ be the time instance at which π takes a patient of class ij for the first time. Below, we consider cases 1 and 2 separately.

Case 1: In this case, the first decision epoch after t_0 is $t_0 + \alpha$. Policy $\hat{\pi}$ follows policy π for $t \in [t_0 + \alpha, \tilde{t})$. At $\tilde{t}$, π chooses a patient from class ij while $\hat{\pi}$ chooses a patient from class $i'j'$, and both policies finish serving them at $\tilde{t} + \beta$, at which the state of the system becomes the same under both policies, and $\hat{\pi}$ follows π from then on. Hence, we have $R_\pi^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t) - R_{\hat{\pi}}^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t) = (\Theta'_{ij}(t) - \Theta'_{ij}(t_0)) - (\Theta'_{i'j'}(t) - \Theta'_{i'j'}(t_0)) = \Theta'_{ij}(t - t_0) - \Theta'_{i'j'}(t - t_0) \geq 0$ for all $t \in [t_0, \tilde{t})$, and $R_\pi^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t) - R_{\hat{\pi}}^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t) = (\Theta'_{ij}(\tilde{t}) - \Theta'_{ij}(t_0)) - (\Theta'_{i'j'}(\tilde{t}) - \Theta'_{i'j'}(t_0)) = \Theta'_{ij}(\tilde{t} - t_0) - \Theta'_{i'j'}(\tilde{t} - t_0) \geq 0$ for all $t \geq \tilde{t}$ (and clearly $R_\pi^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t) - R_{\hat{\pi}}^{\boldsymbol{\mathcal{U}}' \cup \boldsymbol{\mathcal{C}}'}(t) = 0$ for all $t < t_0$).

Case 2: In this case, $\hat{\pi}$ follows π at all decision epochs in $[t_0+m,\tilde{t}+m-M)$ and serves a patient of class $i'j'$ at $\tilde{t}+m-M$ while π serves a patient of class ij at time $\tilde{t}$, the state under π and $\hat{\pi}$ becomes the same at $\tilde{t}+m$, and $\hat{\pi}$ follows π from then on. Thus, we have $R_{\pi}^{\mathcal{U}'\cup\mathcal{C}'}(t)-R_{\hat{\pi}}^{\mathcal{U}'\cup\mathcal{C}'}(t)=(\Theta'_{ij}(t)-\Theta'_{ij}(t_0))-(\Theta'_{i'j'}(t)-\Theta'_{i'j'}(t_0))=\Theta'_{ij}(t-t_0)-\Theta'_{i'j'}(t-t_0)\geq 0$ for all $t\in[t_0,t_0+m)$, $R_{\pi}^{\mathcal{U}'\cup\mathcal{C}'}(t)-R_{\hat{\pi}}^{\mathcal{U}'\cup\mathcal{C}'}(t)\geq(\Theta'_{ij}(t)-\Theta'_{ij}(t_0))-(\Theta'_{i'j'}(t)-\Theta'_{i'j'}(t_0))=\Theta'_{ij}(t-t_0)-\Theta'_{i'j'}(t-t_0)\geq 0$ for all $t\in[t_0+m,\tilde{t}+m-M)$ (where the first inequality holds since all the other patients in this interval are served under $\hat{\pi}$ no later than that under π), $R_{\pi}^{\mathcal{U}'\cup\mathcal{C}'}(t)-R_{\hat{\pi}}^{\mathcal{U}'\cup\mathcal{C}'}(t)\geq(\Theta'_{ij}(t)-\Theta'_{ij}(t_0))-(\Theta'_{i'j'}(\tilde{t}+m-M)-\Theta'_{i'j'}(t_0))\geq(\Theta'_{ij}(t)-\Theta'_{ij}(t_0))-(\Theta'_{i'j'}(t)-\Theta'_{i'j'}(t_0))=\Theta'_{ij}(t-t_0)-\Theta'_{i'j'}(t-t_0)\geq 0$ for all $t\in[\tilde{t}+m-M,\tilde{t})$, and $R_{\pi}^{\mathcal{U}'\cup\mathcal{C}'}(t)-R_{\hat{\pi}}^{\mathcal{U}'\cup\mathcal{C}'}(t)\geq(\Theta'_{ij}(\tilde{t})-\Theta'_{ij}(t_0))-(\Theta'_{i'j'}(\tilde{t}+m-M)-\Theta'_{i'j'}(t_0))\geq(\Theta'_{ij}(\tilde{t})-\Theta'_{ij}(t_0))-(\Theta'_{i'j'}(\tilde{t})-\Theta'_{i'j'}(t_0))=\Theta'_{ij}(\tilde{t}-t_0)-\Theta'_{i'j'}(\tilde{t}-t_0)\geq 0$ for all $t\geq\tilde{t}$ (and clearly $R_{\pi}^{\mathcal{U}'\cup\mathcal{C}'}(t)-R_{\hat{\pi}}^{\mathcal{U}'\cup\mathcal{C}'}(t)=0$ for all $t<t_0$). This completes the proof. □

COROLLARY EC.1 **(General Adverse Events)**. *If the four patient classes in $\mathcal{U}\times\mathcal{C}$ can be relabeled such that $\{\Theta'_{[4]}(t),t\geq 0\}\leq_{st}\{\Theta'_{[3]}(t),t\geq 0\}\leq_{st}\{\Theta'_{[2]}(t),t\geq 0\}\leq_{st}\{\Theta'_{[1]}(t),t\geq 0\}$ and $S'_{[1]}\leq_{lr}S'_{[2]}\leq_{lr}S'_{[3]}\leq_{lr}S'_{[4]}$, then prioritizing patients in the increasing order of their class labels stochastically minimizes $R_{\pi}^{\mathcal{U}'\cup\mathcal{C}'}(t)$ at every $t\geq 0$ for $\pi\in\Xi$.*

Proof of Corollary EC.1: The result follows directly from Proposition EC.1 using a simple interchange argument. □

The above results presents a version of Proposition 2 (i) presented in the main body for the case where the Poisson assumption on the occurrence of adverse events is relaxed. The implication is that a similar priority rule remains optimal, provided that the complexity-based classification is defined in such a way that the "agreeable" ordering conditions of Corollary EC.1 hold.

Online Appendix D (The Clearing Model of Phase 2 ED Service)

Consider the simplified Phase 2 model with dynamic arrivals but, instead of assuming patients *classified* as $ij\in\mathcal{U}\times\mathcal{C}$ arrive according to a Poisson process with rate λ'_{ij} (superscript " $'$ " indicates error-impacted rates), assume there are a finite number of patients already in the ED and available to the physician (and there is no further arrival). The physician wants to clear the system in a manner that minimizes the expected number of adverse events. If the number of patients exceeds the number the provider can treat in parallel, then this scenario approximates backup overload periods.

To derive the optimal physician's policy, we use the same notation used in the manuscript for the simplified Phase 2 model with dynamic arrivals. The following theorem shows that the priority rule presented in Theorem 1 remains optimal.

THEOREM EC.1 **(Phase 2 Prioritization- Clearing)**. *In the simplified Phase 2 model with* *static arrivals, regardless of the number and class of available and unavailable patients, the physician should prioritize available patients in decreasing order of* $p'_{ij}\,\hat{\theta}'_{ij}\,\hat{\mu}'_{ij}$. *Furthermore, the physician should not idle when there is a patient available in an exam room.*

Proof of Theorem EC.1: Let $(\underline{x},\underline{y})$ be the sate of the system, where $\underline{x}=(x_{ij}:ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}})$ is the error-impacted number of the patients "available" to the physican and $\underline{y}=(y_{ij}:ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}})$ is the error-impacted number of the patients "unavailable" (i.e., under test, waiting for the results, etc.). Also let $b<\infty$ denote the number of patients in the system at time zero (note that we have $(\underline{x},\underline{y})\,\underline{1}^T_{1\times 8}\le b$ at all time, where $\underline{1}$ denotes a vector of ones.). The goal is to move the system from a particular state $(\underline{x},\underline{y})$ with $(\underline{x},\underline{y})\,\underline{1}^T_{1\times 8}=b$ to the absorbing state $(\underline{0},\underline{0})$, using a policy that results in the minimum expected number of adverse events. Note that since $(\underline{0},\underline{0})$ is an absorbing state, the problem can be modeled as an infinite-horizon Markov Decision Process (MDP) with a discount factor $\beta\triangleq 1$ and a bounded transition rate $\psi=\max_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}\hat{\mu}'_{ij}+b\,\eta<\infty$.

After uniformization, the Bellman equation for any state in the state space $\mathcal{S}\triangleq\{(\underline{x},\underline{y})\in\mathbb{Z}^8_+:(\underline{x},\underline{y})\,\underline{1}^T_{1\times 8}\le b\}$ can be written as follows (with the boundary condition $V(\underline{0},\underline{0})=0$):

$$\begin{aligned}V(\underline{x},\underline{y})=\frac{1}{\psi}\Big[&\underline{\hat{\theta}'}\,(\underline{x}+\underline{y})^T+\beta\big[\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[y_{ij}\,\eta\,V(\underline{x}+\underline{e}_{ij},\underline{y}-\underline{e}_{ij})]\\&+\min_{a\in\mathcal{A}(\underline{x})}\big\{\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}\mathbb{1}_{\{a=ij\}}\hat{\mu}'_{ij}\,[p'_{ij}\,V(\underline{x}-\underline{e}_{ij},\underline{y})+(1-p'_{ij})V(\underline{x}-\underline{e}_{ij},\underline{y}+\underline{e}_{ij})]\\&+\big(\psi-\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[y_{ij}\,\eta+\mathbb{1}_{\{a=ij\}}\hat{\mu}'_{ij}]\big)\,V(\underline{x},\underline{y})\big\}\big]\Big],\end{aligned}\tag{EC.12}$$

where $\underline{e}_{ij}$ is a vector with the same size as $\underline{x}$ with a 1 in position ij and zeroes elsewhere, a is an action determining which patient class to serve, and $\mathcal{A}(\underline{x})=\{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}:x_{ij}>0\}\cup\{0\}$ is the set of feasible actions (class 0 represents the idling action which is considered as serving an extra class with all the rates equal to zero) when the error-impacted number of patients of each class who are "available" is $\underline{x}$.

To show the result, we use the finite-horizon equivalent optimality equation. When there are $n+1$ (uniformized) periods to go, it can be written as:

$$\begin{aligned}V_{n+1}(\underline{x},\underline{y})=\frac{1}{\psi}\Big[&\underline{\hat{\theta}'}\,(\underline{x}+\underline{y})^T+\beta\big[\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[y_{ij}\,\eta\,V_n(\underline{x}+\underline{e}_{ij},\underline{y}-\underline{e}_{ij})]\\&+\min_{a\in\mathcal{A}(\underline{x})}\big\{\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}\mathbb{1}_{\{a=ij\}}\hat{\mu}'_{ij}\,[p'_{ij}\,V_n(\underline{x}-\underline{e}_{ij},\underline{y})+(1-p'_{ij})V_n(\underline{x}-\underline{e}_{ij},\underline{y}+\underline{e}_{ij})]\\&+\big(\psi-\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[y_{ij}\,\eta+\mathbb{1}_{\{a=ij\}}\hat{\mu}'_{ij}]\big)\,V_n(\underline{x},\underline{y})\big\}\big]\Big].\end{aligned}\tag{EC.13}$$

We then use a convergence argument to prove the result for the original infinite horizon MDP. Grouping the terms related to control in the minimization and self-loop terms in (EC.13), we can re-write it as:

$$V_{n+1}(\underline{x},\underline{y})=\frac{1}{\psi}\Big[\underline{\hat{\theta}'}\,(\underline{x}+\underline{y})^T+\beta\big[\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[y_{ij}\,\eta\,V_n(\underline{x}+\underline{e}_{ij},\underline{y}-\underline{e}_{ij})]$$

$$- \max_{a\in\mathcal{A}(\underline{x})}\Big\{ \sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}} \mathbb{1}_{\{a=ij\}}\hat{\mu}'_{ij}\,[p'_{ij}\Delta^{y}_{ij}\,V_n(\underline{x}-\underline{e}_{ij},\underline{y}) + \Delta^{x,y}_{ij}V_n(\underline{x}-\underline{e}_{ij},\underline{y}+\underline{e}_{ij})]$$
$$+\big(\psi - \sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}} y_{ij}\,\eta\big)\,V_n(\underline{x},\underline{y})\Big\}\Big]\Big], \tag{EC.14}$$

where $\Delta^{y}_{ij}\,V_n(\underline{x},\underline{y}) = V_n(\underline{x},\underline{y}+\underline{e}_{ij}) - V_n(\underline{x},\underline{y})$ and $\Delta^{x,y}_{ij}V_n(\underline{x},\underline{y}) = V_n(\underline{x}+\underline{e}_{ij},\underline{y}-\underline{e}_{ij}) - V_n(\underline{x},\underline{y})$.

To show the result, we use an *interchange* argument; we show that if classes $uc\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}$ and $sl\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}$ are such that $p'_{uc}\,\hat{\theta}'_{uc}\,\hat{\mu}'_{uc} \geq p'_{sl}\,\hat{\theta}'_{sl}\,\hat{\mu}'_{sl}$, then it is (weakly) better to serve class uc than class sl when in state $(\underline{x},\underline{y})$ with $x_{uc}, x_{sl} > 0$. This will also prove that the optimal policy will not idle the physician when there are one or more patients available in the rooms, since idling can be thought of serving an additional class, class 0, with $\hat{\theta}'_0 = \hat{\mu}'_0 = p'_0 = 0$

Let π ($\hat{\pi}$) be the policy that prescribes serving patients of class uc (sl) for every state $(\underline{x},\underline{y})\in\mathcal{S}$ with $x_{uc}, x_{sl} > 0$ and in every period n. Also, let $V^{\pi}_n(\cdot)$ and $V^{\hat{\pi}}_n(\cdot)$ denote the costs (when there are n periods to go) of policies π and $\hat{\pi}$, respectively. From (EC.14), to show that π is (weakly) better than $\hat{\pi}$ in every period, we need to show that the following property holds for every n and every state $(\underline{x},\underline{y})\in\mathcal{S}$ with $x_{uc}, x_{sl} > 0$:

$$\hat{\mu}'_{uc}\,[p'_{uc}\Delta^{y}_{uc}\,V^{\pi}_n(\underline{x}-\underline{e}_{uc},\underline{y}) + \Delta^{x,y}_{uc}V^{\pi}_n(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})]$$
$$\geq \hat{\mu}'_{sl}\,[p'_{sl}\Delta^{y}_{sl}\,V^{\hat{\pi}}_n(\underline{x}-\underline{e}_{sl},\underline{y}) + \Delta^{x,y}_{sl}V^{\hat{\pi}}_n(\underline{x}-\underline{e}_{sl},\underline{y}+\underline{e}_{sl})]. \tag{EC.15}$$

To show property (EC.15), we use induction on n. First, for $n=0$, the property trivially holds since $V^{\pi}_0(\cdot,\cdot) = V^{\hat{\pi}}_0(\cdot,\cdot) = 0$. Next, suppose the property holds for n. We show that it will then also hold for $n+1$. To do so, we need to consider different cases based on the state (i.e., partitions of the state space). First, if $b\geq 4$ consider the case where $x_{uc}, x_{sl}\geq 2$. Using action $a = uc$ (policy π) in both states $(\underline{x}-\underline{e}_{uc},\underline{y})$ and $(\underline{x},\underline{y})$ to compute $V^{\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})$ and $V^{\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y})$ using (EC.14), and subtracting the results we have $\Delta^{y}_{uc}V^{\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y}) =$

$$\frac{1}{\psi_k}\Big[\hat{\theta}'_{uc} + \beta\big[\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[y_{ij}\,\eta\,\Delta^{y}_{uc}V^{\pi}_n(\underline{x}+\underline{e}_{ij}-\underline{e}_{uc},\underline{y}-\underline{e}_{ij})]$$
$$+\eta\,V^{\pi}_n(\underline{x},\underline{y}-\underline{e}_{uc})$$
$$-\hat{\mu}'_{uc}\,[p'_{uc}\Delta^{y}_{uc}\Delta^{y}_{uc}\,V^{\pi}_n(\underline{x}-2\underline{e}_{uc},\underline{y}) + \Delta^{y}_{uc}\Delta^{x,y}_{uc}V^{\pi}_n(\underline{x}-2\underline{e}_{uc},\underline{y}+\underline{e}_{ij})]$$
$$+\big(\psi - \sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}} y_{ij}\,\eta]\big)\,\Delta^{y}_{uc}V^{\pi}_n(\underline{x}-\underline{e}_{uc},\underline{y}) - \eta\,V^{\pi}_n(\underline{x}-\underline{e}_{uc},\underline{y})\big]\Big]. \tag{EC.16}$$

Similarly, we can derive $\Delta^{x,y}_{uc}V^{\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})$ using (EC.14) and action $a = uc$ (policy π) in both states $(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc})$ and $(\underline{x},\underline{y})$ and subtracting the results. Doing so we have $\Delta^{x,y}_{uc}V^{\pi}_{n+1}(\underline{x}-\underline{e}_{uc},\underline{y}+\underline{e}_{uc}) =$

$$\frac{1}{\psi_k}\Big[\beta\big[\sum_{ij\in\boldsymbol{\mathcal{U}}\times\boldsymbol{\mathcal{C}}}[y^{+}_{ij}\,\eta\,\Delta^{x,y}_{uc}V^{\pi}_n(\underline{x}+\underline{e}_{ij}-\underline{e}_{uc},\underline{y}-\underline{e}_{ij}+\underline{e}_{uc})]$$
$$-\eta\,V^{\pi}_n(\underline{x},\underline{y})$$

$$- \hat{\mu}'_{uc} \left[p'_{uc} \Delta^{x,y}_{uc} \Delta^{y}_{uc} V^{\pi}_{n}(\underline{x} - 2\underline{e}_{uc}, \underline{y} + \underline{e}_{uc}) + \Delta^{x,y}_{uc} \Delta^{x,y}_{uc} V^{\pi}_{n}(\underline{x} - 2\underline{e}_{uc}, \underline{y} + \underline{e}_{ij} + \underline{e}_{uc}) \right] + \Big(\psi - \sum_{ij \in \mathcal{U} \times \mathcal{C}} [y^{+}_{ij}\, \eta] \Big) \Delta^{x,y}_{uc} V_n(\underline{x} - \underline{e}_{uc}, \underline{y} + \underline{e}_{uc}) + \eta\, V^{\pi}_{n}(\underline{x} - \underline{e}_{uc}, \underline{y} + \underline{e}_{uc}) \Big] \Big], \tag{EC.17}$$

where $y^{+}_{ij} = y_{ij}$ for all $ij \neq uc \in \mathcal{U} \times \mathcal{C}$, and $y^{+}_{uc} = y_{uc} + 1$. In a similar way, and by using action $a = sl$ (policy $\hat{\pi}$) in (EC.14) quantities $\Delta^{y}_{sl} V^{\hat{\pi}}_{n+1}(\underline{x} - \underline{e}_{sl}, \underline{y})$ and $\Delta^{x,y}_{sl} V^{\hat{\pi}}_{n+1}(\underline{x} - \underline{e}_{sl}, \underline{y} + \underline{e}_{sl})$ can be computed. Next, to check property (EC.15) for $n+1$, multiply (EC.16) by $p'_{uc}\,\hat{\mu}'_{uc}$, and (EC.17) by $\hat{\mu}'_{uc}$ and add up the results. Similarly, multiply $\Delta^{y}_{sl} V^{k,\hat{\pi}}_{n+1}(\underline{x} - \underline{e}_{sl}, \underline{y})$ and $\Delta^{x,y}_{sl} V^{k,\hat{\pi}}_{n+1}(\underline{x} - \underline{e}_{sl}, \underline{y} + \underline{e}_{sl})$ by $p'_{sl}\,\hat{\mu}'_{sl}$ and $\hat{\mu}'_{sl}$, respectively, and add up the results. Next, using the induction hypothesis and that $p'_{uc}\,\hat{\theta}'_{uc}\,\hat{\mu}'_{uc} \geq p'_{sl}\,\hat{\theta}'_{sl}\,\hat{\mu}'_{sl}$, after algebraic simplification it follows that

$$\hat{\mu}'_{uc} \left[p'_{uc} \Delta^{y}_{uc} V^{\pi}_{n+1}(\underline{x} - \underline{e}_{uc}, \underline{y}) + \Delta^{x,y}_{uc} V^{\pi}_{n+1}(\underline{x} - \underline{e}_{uc}, \underline{y} + \underline{e}_{uc}) \right] - \hat{\mu}'_{sl} \left[p'_{sl} \Delta^{y}_{sl} V^{\hat{\pi}}_{n+1}(\underline{x} - \underline{e}_{sl}, \underline{y}) + \Delta^{x,y}_{sl} V^{\hat{\pi}}_{n+1}(\underline{x} - \underline{e}_{sl}, \underline{y} + \underline{e}_{sl}) \right] \geq 0, \tag{EC.18}$$

which establishes property (EC.15) for $n+1$ for the case where $x_{uc}, x_{sl} \geq 2$. In a similar way, this property can be established for other cases (i.e., the remaining partition of the state space). Hence, the non-idling strict priority rule is optimal for all n. Next, taking the limit as $n \to \infty$ it follows that the finite horizon problem converges to the original infinite horizon one both in policy and cost (see, e.g., Sennott (1999) Proposition 4.3.1), which completes the proof. □

Online Appendix E (Sensitivity Analysis on the Percentage Drop in Phase 2 Intensity of Adverse Events)

We assumed that the intensity of adverse events decreases by 60% when patients move from the waiting room (Phase 1) to an examination room (Phase 2). As we discussed the 60% number is a physician estimate based on the impact of more careful monitoring and care within the ED. In this online appendix, we briefly present sensitivity analysis on this percentage drop in Phase 2 intensity of adverse events to show that our main conclusions (in particular Observation 1) are robust to this estimation.

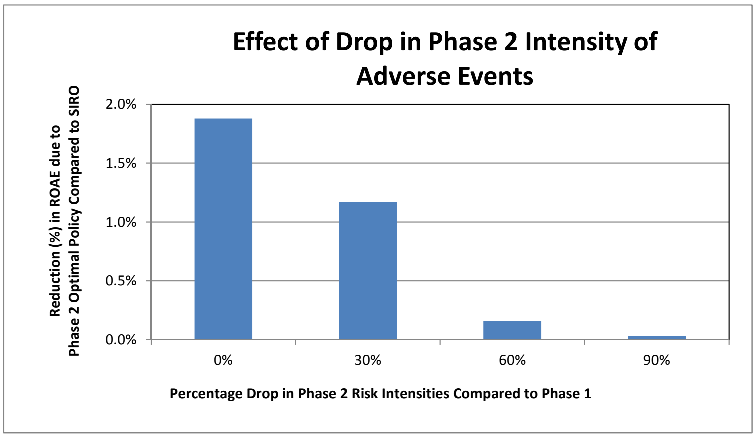

Figure EC.2 **Robustness to the estimated percentage drop in Phase 2 intensity of adverse events.**

Figure EC.2 presents a numerical test that compares ROAE under SIRO and under the optimal Phase 2 policy. This shows that the impact of Phase 2 prioritization is substantially smaller than the impact of Phase 1 sequencing (which we showed was 7.9% in our base case) regardless of the amount by which ROAE decreases from Phase 1 to Phase 2. Even when there is no drop at all, which maximizes the importance of Phase 2 prioritization, the impact of Phase 2 prioritization is less than 2%, which is less than 25% of the impact of Phase 1 sequencing.

References for the Online Appendices

Buyukkoc, C., P. Varaiya, J. Walrand. 1985. The $c\mu$ rule revisited. *Adv. Appl. Prob.* 17, 237–238.

Cox, D.R., W.L. Smith. 1961. *Queues.* Methuen & Co, London.

Guo, X., K. Liu. 2001. A note on optimality conditions for continuous-time Markov Decision Processes with aevrage cost criterion. *IEEE Trans. on Aut. Contr.* 46(12) 1984–1989.

Kakalik, J.S., J.D.C. Little. 1971. *Optimal Service Policy for the M/G/1 Queue with Multiple Classes of Arrival.* Rand Corporation Report.

Righter, R., *Scheduling, "Stochastic Orders and Their Applications",* Chapter 13, edited by M. Shaked and J. G. Shanthikumar. New York: Academic Press, 381–432, 1994.

Sennott, L.I. 1999. *Stochastic Dynamic Programming and the Control of Queueing Systems.* Wiley Series in Probability and Statistics, John Wiley and Sons, New York.

Vance, J., P. Spirvulis. 2005. Triage nurses validly and reliably estimate Emergency Department patient complexity. *Emergency Medicine Australasia* 17, 382–386.