Dynamic capacity allocation for elective surgeries: reducing urgency-weighted wait times

Abstract

Problem definition: Given the variety of urgency levels in highly utilized operating rooms, capacity allocation decisions can have a major impact on how wait times are rationed. We examine a longer-term, sequential capacity planning problem, in which a hospital allocates operating room time to different surgical specialties. We seek to minimize an urgency-weighted wait time metric.

Academic/Practical relevance: Our dataset on patient selection patterns revealed considerable noise in the queueing discipline. We apply an urn model to generate a probabilistic queueing discipline, which validates well against the selection patterns observed in practice. We believe this model may prove useful for representing noisy queueing disciplines in other settings. Also, our validated simulation model, in combination with our proposed solution approach, demonstrates a substantial reduction in urgency-weighed wait times.

Methodology: For representing the noisy queueing discipline, we fit a Wallenius noncentral hypergeometric distribution. We formulate the capacity allocation problem as a Markov decision process. The large state space and detailed system dynamics lead us to simulation-based dynamic programming approaches for finding good capacity allocation decisions. Rather than approximate the expected cost-to-go function, we propose a limited lookahead policy, and embed this in a rolling horizon framework.

Results: Our baseline model-based allocation policy yields a 14.3% reduction in urgency-weighed wait time compared to current practice. It also results in a 21.0% improvement in the number of patients treated within their urgency-based recommended wait time limits.

Managerial implications: In elective surgery settings, it may be important to ration capacity in a way that considers the different urgency levels of patients. We propose a flexible modeling approach for achieving this.