ABSTRACT

We develop a discrete event simulation model for a network of eight major intensive care units (ICUs) in British Columbia, Canada. The model also contains high acuity units (HAUs) that provide critical care to patients that cannot be cared for in a general medical ward, but do not require the full spectrum of care available in an ICU. We model patient flow within the ICU and HAU for each of the hospitals, as well as patient transfers to address ICU capacity. Included in the model is early discharge from ICU to HAU, sometimes called “bumping”, when the ICU is full, as well as ICU overflow beds. The simulation model, which is calibrated using the British Columbia Critical Care Database, will be used to support planning for critical care capacity under endemic and seasonal COVID-19.

1 INTRODUCTION

The COVID-19 pandemic highlighted the vulnerability of the critical care system in many countries to a surge in cases arising from a pandemic or other public health crisis. This project developed from simulation modelling done on utilization of mechanical ventilators in British Columbia, Canada, during the first wave of the COVID-19 pandemic in April to June, 2020 (Zimmerman, Samantha L. and Rutherford, Alexander R. and van der Waall, Alexa and Norena, Monica and Dodek, Peter 2022). With the likelihood that ongoing endemic COVID-19 and seasonal outbreaks will continue to place burdens on the critical care system, it became apparent that this simulation modelling project should be expanded to a multi-hospital model that incorporates intensive care unit (ICU) beds, staffing limitations, and transfers between hospitals.

This paper describes a multi-hospital simulation model of the critical care network in British Columbia, Canada to support management of the critical system under endemic and seasonal COVID-19. An important aspect of the project is to model the network of major ICUs in the province, reflecting the extent to which
ICUs operated together during the pandemic. Critical care capacity in British Columbia is primarily limited by the availability of medical staff, such as ICU nurses, intensivists, respiratory therapists, and perfusionists. Additional critical care beds were added during the pandemic; however, there were considerable challenges in staffing this extra capacity. To make our simulation model applicable to capacity planning under surge conditions, we incorporated ICU nurses as one of the limiting resources in our model. Furthermore, during the pandemic high acuity units (HAUs) were used to expand capacity of the critical care system and we have included HAUs in our simulation model. HAUs provide care to patients who require closer monitoring than is available in a general hospital ward, but do not require the full spectrum of specialized care available in an ICU.

2 LITERATURE REVIEW

Critical care patients generally require immediate admission to an ICU. However, ICUs are expensive resources, which must be managed carefully to balance cost with patient access. Operations research and simulation modelling have been used to determine how to best manage capacity and patient flow in the ICU (Dong et al. 2012; Bai et al. 2018). Discrete event simulation (DES) has been used to determine ICU bed capacities required to balance patient access and cost effectiveness (Ridge et al. 1998; Costa et al. 2003; Mallor and Azcárate 2014; Zhu et al. 2012; Steins and Walther 2013; Alban et al. 2020). García-Vicuña et al. (2019) have developed a management flight simulator to train ICU staff in making patient flow and capacity decisions in real time.

Early in the COVID-19 pandemic, it was recognized that the pandemic would place considerable strain on the critical care system. Models of COVID-19 case projections were linked to simulation models of the critical care system to project demand for ICU beds and other critical care resources (Currie et al. 2020; Meares and Jones 2020; Zimmerman, Samantha L. and Rutherford, Alexander R. and van der Waall, Alexa and Norena, Monica and Dodek, Peter 2022; Garcia-Vicuña et al. 2022). Discrete event simulation has been used to determine the number of ICU beds required to manage a given number of COVID-19 cases (Alban et al. 2020; Meares and Jones 2020; Lu et al. 2021). Some studies utilized approximation methods from queuing theory to obtain heuristic approaches to estimating ICU utilization from COVID-19 case data (Bekker et al. 2023; Baas et al. 2021). Both the simulation and queuing theory studies helped to motivate efforts to use public health measures to “flatten the curve” of cases. Wood et al. (2020) used simulation modelling of hospital ICUs to project COVID-19 deaths due to hospital capacity limitations during the first wave of the pandemic. DES was also used to examine the policy of reducing elective surgeries during the pandemic to reduce the demand for ICU beds and staff (Melman et al. 2021).

Strategies employed by critical care staff to admit patients when the ICU is full include using overflow beds elsewhere in the hospital, discharging patients near the end of their stay to HAUs or medical wards, or transferring patients to ICUs in other hospitals (Chan et al. 2018). Steins and Walther (2013) used data from four ICUs in Sweden to develop a detailed simulation model that captures overflow beds and patient transfers. The process of discharging patients near the end of their stay to make beds available for new critical care admissions is sometimes referred to as “bumping” (Reader et al. 2018). Simulation modelling has been used to evaluate different policies for early discharge and use of overflow beds to manage patient access when the ICU is at or near capacity (Mallor and Azcárate 2014; Bai et al. 2020). Dobson et al. (2010) developed a Markov chain model for the ICU bumping process. Queuing models have also been used to relate key performance indicators (KPIs) for patient care to different strategies for managing patient flow in the ICU (Hagen et al. 2013; Varney et al. 2019). Discrete event simulation models have been used to relate ICU admission and discharge policies to KPIs (Hasan et al. 2020; Bahalkeh et al. 2022).

HAUs or “step-down” units provide care for critical care patients who do not require care in the ICU, but cannot be cared for in a general medical ward (Mitra et al. 2020). These units are an important part of the critical care system and their use has been increasing in recent years to relieve pressure on ICUs. During the COVID-19 pandemic, HAUs provided an important part of the critical care surge capacity. Rodrigues et al. (2018) used a DES model to demonstrate the benefits of using a “step-down” unit to improve patient
flow in the ICU. During the pandemic in British Columbia, HAUs are not only used as “step-down” units for patients discharged from the ICU, but patients were admitted directly to the HAU. A recent trend in the management of critical care is to consider the ICU and HAU as an integrated system (Mitra et al. 2020).

3 DATA ANALYSIS

The primary source of data for this project is an extract from the British Columbia Critical Care Database for years 2010–20. This database contains detailed data on admissions to ICUs and HAUs at most hospitals in British Columbia. This includes data on the patient treatment in the ICU, time of patient admission and discharge, patient transfers, start and stop times for all mechanical ventilation instances, and diagnostic information on the patient. This database was initiated in 1997 as the British Columbia ICU Database. We linked the Critical Care Database to the Discharge Abstract Database (DAD) for British Columbia to account for missing data in the Critical Care Database for some hospitals. The DAD contains complete data on all hospitals admissions; however, it contains much less detailed information on ICUs and HAUs. We only considered general ICUs in our data analysis, and not specialized ICUs such as cardiac surgery ICUs. Calibration and validation of our model used primarily 2019 data, because 2019 is the most recent complete year of data prior to the COVID-19 pandemic. Pressures on the critical care system during the pandemic created challenges for data completeness and interpreting data in the context of frequently changing temporary capacity. Therefore, we decided not to use 2020 data for calibrating and validating our model.

The arrival rates for each of the ICUs in the simulation model were calibrated from the patient admission data in the Critical Care Database. Arrival rates were analyzed both by time of day and day of week, and the results are shown in Figure 1. Although the simulation model contains only the eight major ICUs in British Columbia, arrival rates were calculated for each of the 23 hospitals in the Critical Care Database. We found only slight variation in the admission rates by day of week, but there is a significant variation by time of day. Therefore, the non-homogeneous Poisson processes that simulate patient arrivals in the model were calibrated to use a different arrival by hour of day, but day of week dependence was not included in the model.

Patients who received mechanical ventilation would sometimes be placed on a ventilator multiple times. Figure 2 shows the mean and standard deviation of the number of instances on mechanical ventilation and the time on mechanical ventilation for trauma, surgical, and medical patients. These categories of patients are further subdivided into patients with either viral pneumonia (VP) or acute respiratory distress syndrome (ARDS), and patients with neither VP nor ARDS. There were only a small number of trauma and surgical patients with VP or ARDS, which is reflected in the large standard deviations for the number of mechanical ventilation instances and the mean time on mechanical ventilation for these two categories.

The Critical Care Database records the avoidable time in the ICU for each patient, which is defined as the time from when the patient is clinically ready for discharge until the patient is actually discharged from the ICU. Avoidable time usually occurs when a bed is unavailable in either the HAU or the general medical ward, and the patient is unable to be discharged immediately from the ICU. The avoidable time distribution for hospital 1 is shown in Figure 3. The distribution of avoidable times for all of the ICUs has a sequence of peaks which reflect standard times during hospital operations when patients are usually moved from the ICU to either the HAU or medical ward. We fit the avoidable time distribution using a sequence of truncated normal distributions. The distribution of avoidable for each of the ICUs in the model is similar; although, different distribution fits are done for each ICU.

Official bed counts for each of the ICUs were available from the British Columbia Health Authority Management Information System (HAMIS). However, we did not have data on the number of overflow ICU beds available at each of the hospitals. In reality, this is not a fixed number and depends on the ability of the ICU to respond to an exigency in the moment. However, when analyzing ICU occupancy for each of the hospitals we noticed a consistent ceiling on the number of admitted critical care patients. This was
Figure 1: Arrival rate by hours of day and day of week for ICUs at 23 hospitals in British Columbia. Each colour represents a different hospital.

Figure 2: The mean number of mechanical ventilation instances and mean time on mechanical ventilation in hours for patients who received mechanical ventilation. The lines show the standard deviation. VA denotes patients with either VP or ARDS, and NVA denotes patients with neither VP nor ARDS.
Figure 3: Avoidable time at the ICU in hospital 1. The blue curve is the fitted distribution.

used to estimate the number of overflow beds available at the hospital for the purpose of determining when transfers would be necessary.

The Critical Care Database does not record the time patients spend in the HAU and these data were obtained from the DAD. A separate gamma distribution fit for the LOS in the HAU was done for each of the large ICUs in the model. A single gamma distribution was done collectively for the smaller ICUs in the model to ameliorate a problem with missing data for these hospitals.

4 SIMULATION MODEL

A discrete event simulation model of a network ICUs and HAUs in eight hospitals was developed. The development of the model was based on data from the British Columbia Critical Care Database, which captures patient admissions, as well as details of patient flow and treatment, for most hospitals in British Columbia. To understand patient flow and operational practice within the ICUs, two of the authors (ARR and SLZ) shadowed rounds at a major ICU with one of the authors (DEGG). Hospitals in the network contain both ICUs and HAUs. The simulation model is implemented in AnyLogic and a diagram of the model is shown in Figure 4. Multiple hospitals in the AnyLogic model are implemented using AnyLogic resource pools. Each service node contains simulated patients for all hospitals in the modelled critical care network, but the simulated patient only accesses resources for the hospital at which they are located.

Although the Critical Care Database contains detailed information on patient diagnosis, the simulation model considers only four categories of patient diagnoses that are most indicative of LOS in the ICU and whether mechanical ventilation is required. The four categories are: (1) patients diagnosed with either VP or ARDS; (2) medical patients with neither VP nor ARDS; (3) surgical patients with neither VP nor ARDS; and (4) trauma patients with neither VP nor ARDS. Medical, surgical, or trauma patients with VP or ARDS were considered together as one category, because most of the ICUs had small patient volumes in the subcategories. The APACHE II fields in the Critical Care Database were used to classify patients into these categories. Arrival rates and service time distributions in the simulation were calibrated separately for each of these categories of patients.

The simulation model contains six types of service time distributions. For patients who receive mechanical ventilation, probability distributions were calibrated for the time before mechanical ventilation, frequency of ventilation instances, duration of each ventilation instance, time between ventilation instances, and time after mechanical ventilation until being discharged from the ICU. A length of stay in the ICU
distribution is calibrated for patients who do not receive mechanical ventilation. Based on expert opinion, we include in the model a deterministic 2-hour cleaning and servicing time for mechanical ventilators after each use.

Three resources within ICUs and HAUs are tracked by the simulation: beds, mechanical ventilators, and ICU nurses. The nurse to patient ratio is 1:1 in an ICU and 1:2 in an HAU. Patients who require mechanical ventilation are placed on a mechanical ventilator if one is available. The mechanical ventilation component of the model is a generalized Erlang loss queuing model based on the model developed by Zimmerman, Samantha L. and Rutherford, Alexander R. and van der Waall, Alexa and Norena, Monica and Dodek, Peter (2022). In this model, patients are lost to the system if a ventilator is unavailable when required. Mechanical ventilation does not necessarily occur as a single contiguous time within the ICU, but patients may have multiple instances of mechanical ventilation within their ICU stay. When a patient is not using a mechanical ventilator, it is available for use by another patient. Mechanical ventilation is only available within the ICU; patients cannot receive mechanical ventilation in the HAU. Beds are not always treated as a hard resource constraint in the ICU. Patients in the ICU may be moved to the HAU (bumped) to make a bed available for a new admissions, or patients may be placed in overflow ICU beds, provided that nursing staff is available. Each ICU in the model has a fixed number of overflow beds available.

In reality, the decision to bump a patient to the HAU in order to make a bed available for a new admission is based on a complex clinical assessment. In the simulation model, we use a simplified proxy for this decision. Specifically, simulated patients may be bumped if they have completed 85% of their course of treatment, obtained by sampling a projected length of stay (LOS) probability distribution, and they are not using a ventilator. Patients are only bumped from the ICU to the HAU, and not to an overflow bed or a general medical ward bed. The projected LOS does not change when a patient is bumped and the patients complete their ICU LOS in the HAU, prior to beginning their projected LOS for the HAU.

For each of the hospitals in the network, simulated patient arrivals are generated using a non-homogeneous (time-varying) Poisson process. The simulation model first attempts to admit the simulated patient to the “home” ICU, but the patient is transferred to another ICU if there is insufficient capacity. The detailed algorithm for determining if there is sufficient capacity to admit a patient is as follows:

1. If no nurses in either the ICU or HAU are available, then the patient is transferred to another ICU, regardless of bed availability.
2. If there is a nurse in the HAU that is caring for only 1 patient, there is a patient in the ICU that can be bumped, and there is a free HAU bed, then an ICU patient is bumped to the HAU and the arriving patient is admitted to the ICU.
3. If there is at least one fully available nurse:
   (a) If there are ICU beds available then the patient is admitted.
   (b) If there are no ICU beds available:
      i. If there are HAU beds available and a patient can be bumped, then the patient will be bumped to make an ICU bed available.
      ii. If there are no patients that can be bumped or an HAU bed is unavailable:
         ●If an overflow bed is available, then the patient is admitted into an overflow bed.
         ●If there is transfer capacity, then the patient is transferred.
         ●Otherwise the patient is lost to the system.

The simulation model only captures patient transfers due to capacity limitations. Currently, the model does not transfer patients for medical reasons. The model contains a transfer priority matrix that gives for each ICU the priority list of destination ICUs for transfers.

The simulation model records key performance indicators (KPIs) for resource utilization and patient outcomes. The primary KPI for resource utilization is the expected number of occupied ICU or overflow beds at each ICU. For patient outcomes, the simulation model records the proportion of patients at each
Figure 4: Diagram of the AnyLogic model for all ICUs and HAUs in the critical care network. Units at individual hospitals are implemented within the simulation model by allocating hospital-specific resources in AnyLogic.

ICU who are unable to access a ventilator when needed, the proportion of patients at each ICU that are bumped to the HAU due to the ICU reaching capacity, and the proportion of patients at each ICU who are transferred.

5 MODEL VALIDATION

Model validation was done by running the model calibrated with 2019 data and comparing it to mean ICU occupancy for 2019 from the Critical Care Database for each of the eight hospitals. ICU occupancy was calculated from the Critical Care Database by tracking admissions and discharges over multiple years. The simulation was run for 406,944 hours of simulation time and the state of the system was recorded every 169 hours, which is one week plus one hour. This provided approximately independent measurements of the state of the system, which stepped through the week one hour at a time. The result was 2,400 data points on the state of the system, or 100 data points for each hour of the day.

Table 1 shows the comparison of the simulation results with the mean occupancy obtained from the Critical Care Database. Confidentiality precludes us from identifying the hospitals by name, and they are simply numbered one through eight. Hospitals 1, 4, and 5 have large ICUs that provide specialized care. Although hospital 8 does not have a large ICU, it is a major regional centre for critical care. The simulation results for mean ICU occupancy agree with the data from the Critical Care Database to within approximately 4%. The general trend is that the simulation underestimates occupancy for the larger ICUs in the network. We conjecture that this discrepancy occurs because our simulation model only captures patient

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Mean ICU Occupancy 2019 Data</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31.8</td>
<td>30.4</td>
</tr>
<tr>
<td>2</td>
<td>16.8</td>
<td>17.0</td>
</tr>
<tr>
<td>3</td>
<td>8.9</td>
<td>9.3</td>
</tr>
<tr>
<td>4</td>
<td>15.2</td>
<td>14.4</td>
</tr>
<tr>
<td>5</td>
<td>18.5</td>
<td>17.5</td>
</tr>
<tr>
<td>6</td>
<td>9.1</td>
<td>9.0</td>
</tr>
<tr>
<td>7</td>
<td>7.8</td>
<td>8.1</td>
</tr>
<tr>
<td>8</td>
<td>9.5</td>
<td>8.2</td>
</tr>
</tbody>
</table>
transfers for capacity-related reasons, and not medical reasons. Patients are sometimes transferred to larger ICUs, because they require a specialized care or a higher level of care than is available at a smaller ICU. Furthermore, hospital 1 is the major quaternary care center and its ICU always accepts patients requiring specialized care, regardless of capacity limits.

6 DISCUSSION
In British Columbia, the critical care system operated at or above capacity prior to the COVID-19 pandemic, making it difficult for the system to deal with surges in demand during the pandemic. This experience highlighted the need for operational planning to prepare for such surges in critical care demand. Although it is challenging to predict future COVID-19 waves, we will use this simulation model of the critical care network in British Columbia to support planning under a variety of different scenarios being developed by the British Columbia Centre for Disease Control (BCCDC). We aim to use this simulation model to develop strategies for managing the combined impacts of COVID-19 and seasonal influenza without the need for extensive public health interventions to limit transmission.

Increasingly, HAUs are playing an important role in the critical care system. HAUs have the potential to provide a cost-efficient approach to increasing capacity during normal operations, and also provide surge capacity for a crisis. We plan to use our simulation model to optimize policies for integrating HAUs into the critical care system. Important questions include the optimal size of HAUs, and how best to utilize them in the patient pathways as step-down units, step-up units, and units which provide the full duration of critical care.

During the pandemic additional critical care beds were added in many hospitals. Staffing these beds placed considerable pressure on the ICU medical staff, often requiring ICU nurses to work extensive overtime (Crowe et al. 2022). The importance of incorporating medical staff in our simulation model was identified during our consultations with critical care experts in British Columbia. It was felt that this would make our model more useful for future planning of contingencies where ICU nurses and other medical staff are temporarily transferred from other hospitals to address surges in critical care demand at a particular hospital within the network. Working with the BCCDC, we will use this model to determine early warning signals for both COVID-19 and seasonal influenza that could trigger preparation for these contingency measures. Future development of our simulation model will include additional specialized staff, such as respiratory therapists who support mechanical ventilation and perfusionists who support extracorporeal membrane oxygenation (ECMO).

Most previous simulation modelling for critical care has focused on modelling single ICUs, or single hospitals. By modelling a network of ICUs, we are able to study scenarios in which surges in demand at a particular hospital, or within a single region, can be supported by other hospitals in the network. Although the pandemic was a worldwide event, in British Columbia peaks in COVID cases did not occur uniformly across the province. Our model will be used to evaluate the ability of the critical care system to respond as an integrated network by having ICUs outside of the outbreak region provide additional capacity to support ICUs directly impacted by the outbreak. Currently our simulation model does not capture patient transfer resources, such as ambulances or air ambulances. However, patient transfer resources could be added to future versions of the model.

Incorporating transfers for medical reasons into the model is complex, because it would require that the simulation capture aspects of medical decision-making. Machine learning has been used to analyze critical care data for this purpose (Shillan et al. 2019). Recently, machine learning has been used to predict COVID-19 severity from testing data and other patient characteristics (Khoshbakhtian, Faraz and Lagman, Ardian and Aleman, Dionne M. and Giffen, Randy and Rahman, Proton 2021; Patel et al. 2021). Although we do not envision using machine learning to accurately predict outcomes for individual patients, we are currently working on combining machine learning with our DES model to develop a hybrid model that takes into account transfers for medical reasons. This would allow the model to address access to specialized treatments, such as ECMO, which may be provided at only major ICUs within the network.
ACKNOWLEDGMENTS

This project was supported in part by a Health Systems Improvement grant from the British Columbia (BC) Ministry of Health. We thank Sandra Feltham and Quiya Xu at the BC Ministry of Health for supporting this project. ARR and SLZ thank Vancouver General Hospital for the opportunity to shadow ICU rounds and gain insight into patient flow within the ICU. We also thank the Critical Care Services Executive Committee of British Columbia for providing access to the Critical Care Database and expert advice on the critical care network in BC.

REFERENCES


AUTHOR BIOGRAPHIES

ALEXANDER R. RUTHERFORD is the Scientific Director of the Centre for Operations Research and Decision Science and an Adjunct Professor in the Department of Mathematics at Simon Fraser University. His research interests include healthcare operations research, public health, and disease modelling. His e-mail address is arruther@sfu.ca.

SAMANTHA L. ZIMMERMAN is a recent graduate with a Masters in Science from the Mathematics Department at Simon Fraser University. Her research focus and interests include healthcare operations research, queuing theory, simulation, and discrete optimization. Her email address is samantha.zimmerman@sfu.ca.

MINA MOEINI is a second-year graduate student and Research Assistant in Applied mathematics at Simon Fraser University. Her research is focused on operations research, mathematical disease modelling and applications of data-driven techniques for decision supporting real-world healthcare problems. Her email address is mina_moeini@sfu.ca.

RASHID BARKET is a research assistant at the Centre for Operations Research and Decision Science and the SFU Big Data Hub. He has completed the Mathematics and Computing Science Joint-Major undergraduate degree at Simon Fraser University. His research interests are Machine Learning and Operations Research. His email address is rashid_barket@sfu.ca.

STEVE AHKIOON is a critical care data advisor at Vancouver Coastal Health and based at Vancouver General Hospital. His interest is in helping to leverage healthcare data to provide evidence-driven information to support healthcare planning, operations, quality and safety to enhance patient outcomes/experience. His email address is Steve.Ahkioon@vch.ca.

DONALD E. G. GRIESDALE is an Assistant Professor in the Division of Critical Care Medicine, Faculty of Medicine, University of British Columbia. His research interests are learning health systems for critical care, neurocritical care, traumatic brain injury, invasive neuromonitoring, and airway management. His email address is Donald.Griesdale@vch.ca.