



A Monte Carlo simulation-based simulated annealing algorithm for predicting the minimum staffing requirement at a blood donor centre

Achini Wellalage¹ · Mark Fackrell¹ · Lele Zhang¹

Accepted: 14 March 2023

© The Author(s) 2023

Abstract

Australian Red Cross Lifeblood collects blood from non-remunerated voluntary donors. Thus, it is important to ensure that donors experience good service so they will return to donate blood again. Donor experience is adversely influenced by prolonged waiting times, but they may be reduced by determining the staffing demand over the day. In this paper, we propose a Monte-Carlo simulation-based simulated annealing algorithm that seeks the minimum number of employees to meet demand over a single day while ensuring the system's predicted average waiting time does not exceed a specified threshold. To enhance the efficiency of our simulated annealing algorithm, we develop a novel neighbourhood search method based on the staff occupancy levels. We use data from four different Australian Red Cross Lifeblood donor centres, demonstrating that our methodology can be adapted to any donor centre to determine the minimum staffing demand. Since these staffing demands ensure the donor waiting time target is met for each donor centre, they have the potential to improve both donor and staff satisfaction as well as streamline the donor flow.

Keywords Staffing demand · Blood donor centre · Simulation · Simulated annealing · Neighbourhood search · Stochastic optimisation

1 Introduction

Australian Red Cross Lifeblood (ARCL) is a non-profit health service provider funded by the Australian governments. It mainly provides blood or blood products while supplying breast

Mark Fackrell and Lele Zhang have contributed equally to this work.

✉ Achini Wellalage
aerandi@student.unimelb.edu.au

Mark Fackrell
fackrell@unimelb.edu.au

Lele Zhang
lele.zhang@unimelb.edu.au

¹ School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia

milk, transplantation, and biological products for world-class health outcomes (ARCL, 2022). They have collected blood from more than 500,000 donors at 96 static and mobile donor centres across Australia during 2020–2021. In this research, we use data from four different Australian Red Cross Lifeblood donor centres in Melbourne CBD, Gosford, Townsville and Warrnambool to determine their minimum staffing requirements.

For each of the four donor centres, the donation processes are very similar. Donors who arrive at a donor centre go through a rigorous donation process to complete the donation. A donor can make an appointment before arriving at the donor centre. However, scheduled donors may fail to show up or be unpunctual for their appointments. Unscheduled donors can be accommodated, if there is enough capacity and machines available and if there are not already too many donors waiting to be seen. Otherwise, the donor is booked for the next available appointment.

The donation process is shown in Fig. 1. There are three key stages: registration, assessment, and blood collection. Every donor first registers at the reception desk and completes an electronic health questionnaire provided by the receptionist. Next, the donor proceeds to the assessment stage, where a nurse checks the donor's eligibility to donate blood and administers some simple health tests, such as hemoglobin tests, in an interview room. The final key stage, blood collection, consists of three main steps. First, a nurse performs the needle-in, and then a machine that is specialised in collecting whole blood or blood components such as plasma or platelets collects the blood or component. Nurses can assist other donors while a machine collects blood. When the collection is complete, the machine automatically stops, and an available nurse conducts the needle-out. Donors should then rest for a few minutes on the same couch and at least fifteen minutes in the refreshment area before leaving the donor centre. Donors may have to queue for registration, assessment, and needle-out (waiting while lying on the couch after the machine stops drawing blood) due to the unavailability of staff, as depicted in Fig. 1. A donor, in particular, may have to wait until a staff member, couch, and appropriate machine are available to start the needle-in. The first-in-first-out policy applies to all queues, but it can be violated in the needle-in queue since priority is given to the very first donor in the queue whose collection type matches an available machine.

Blood collection in Australia entirely depends on voluntary blood donors. The donations from regular donors are safer and more economical than those from new donors; hence, repeating donations is encouraged (Devine et al., 2007). Furthermore, recommendations from existing donors influence third parties to become new donors. In Australia, only 57.5% of first-time donors make a subsequent donation within 12 months, and the return rates are affected mainly by donors' satisfaction with their previous visit (Masser et al., 2016). Therefore, it is important to provide a quality service that will encourage repeat donations. Donor satisfaction relies on the time spent inside the donor centre. Prolonged waiting times in the queues and long assessment times may adversely affect the donor experience. Therefore, reducing waiting times is crucial to improving donor satisfaction.

ARCL has a *Key Performance Indicator* (KPI) to limit the *Time-to-Couch* (TTC) (that is, the duration from donor arrival to needle-in) to 25 min. According to Fig. 1, TTC is directly affected by the waiting times at the queues for registration, assessment, and needle-in. Therefore, limiting the average waiting time at these three queues to a suitable threshold will accomplish the KPI. We define this threshold as the 'waiting time target'. It can be derived by subtracting the expected times¹ for registration (2 min), questionnaire completion (5 min), and assessment (12 min) from 25 min. Hence, the waiting time target is set at 6 min.

¹ The expected times for registration, questionnaire completion, and assessment were supplied by ARCL.

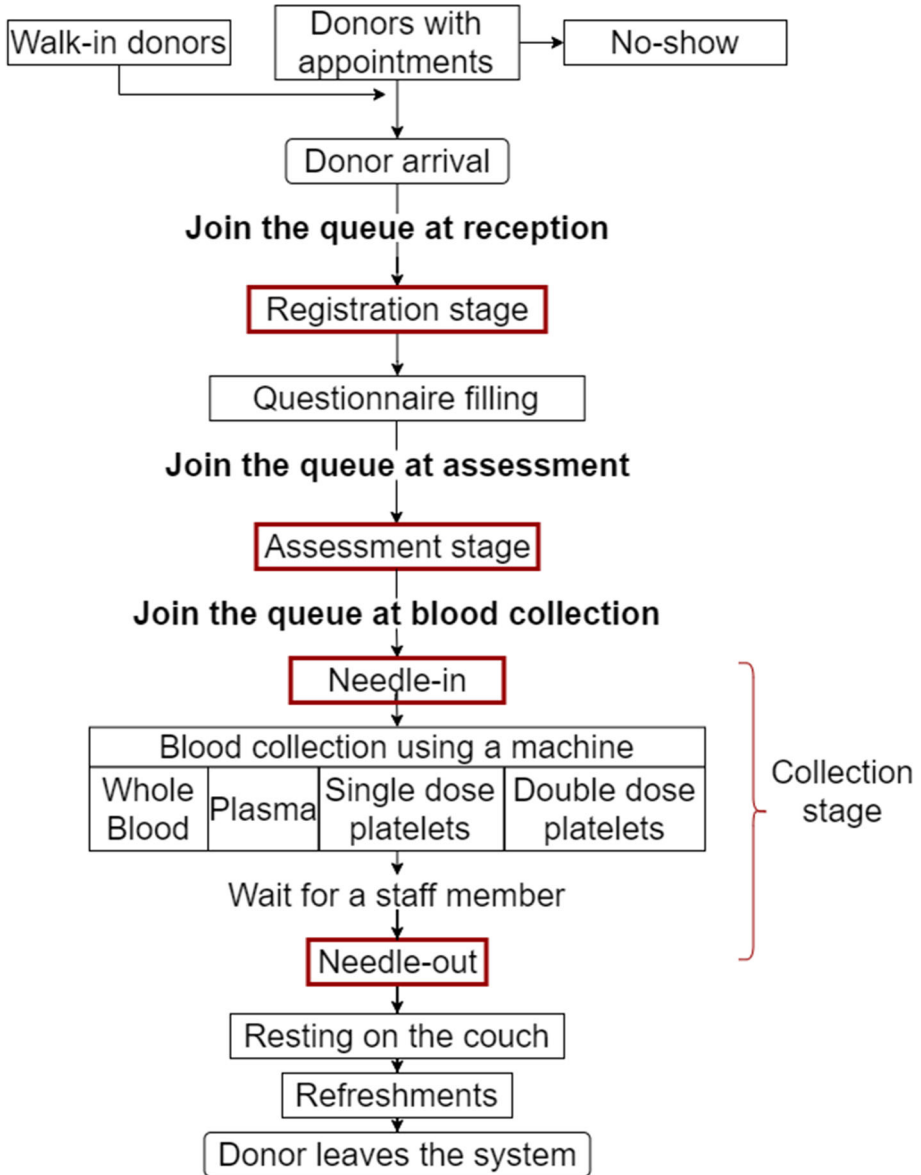


Fig. 1 Donation process

To determine waiting times, modelling the donor flow is essential. We identified three major challenges in modelling the donor flow. The first challenge is modelling donor arrival patterns, which vary due to the earliness and tardiness of appointment-based arrivals, unscheduled arrivals, and no-shows. Second, there are resource limitations to be considered. The facilities that a donor centre can provide are limited, even if the demand for them increases. In a donor centre, there are limited human resources such as receptionists and nurses, as well as physical resources such as reception desk space, interview rooms, blood collection

machines, and couches. Third, uncertainties in processing times due to factors such as staff member skills, donor deferrals for health and safety reasons, and donor adverse events during collection. The distributions of data on activities, including registration, assessment, and collection, are highly positively skewed as a result of these uncertainties. Therefore, we need to model the donation process by considering all sources of uncertainty.

Optimal staffing is a key factor in streamlining donor flow and reducing waiting times. Adding more staff to a donor centre is a simple solution, but it has a negative financial impact as it increases the total number of staff working hours. Therefore, the staffing demand must be minimised. Furthermore, the staffing demand throughout the day is not consistent. This prompts the question, “How do we predict the minimum staffing requirement, which is throughout the day, while keeping the average waiting time at or below its target?”.

In this paper, we propose a stochastic optimisation method to get optimal/near-optimal results for our research question. First and foremost, we fit probability distributions to the data on the duration of registration, questionnaire completion, assessment, and blood collection for all four donor centres. We test a variety of probability distributions by considering the shape of the histograms and the descriptive statistics of the data. The realisations randomly generated from the fitted distributions are then incorporated to develop a comprehensive discrete event simulation model that effectively represents the donation process of a blood donor centre. The simulation model generates accurate and reproducible waiting times so that we can determine the average waiting time associated with a certain staff configuration. Next, we develop an integrated simulation and simulated annealing algorithm that seeks the minimum number of employees to meet varying demands over a single day while ensuring the system’s predicted average waiting time does not exceed the waiting time target. To enhance the efficiency of our simulated annealing algorithm, we develop a novel neighbourhood search method based on the staff occupancy levels.

This paper is organised as follows: First, we give a literature survey of fitting probability distributions to data, simulating a complex queueing system, and minimum staffing problem in Sect. 2. We introduce data we received from the ARCL relevant to four donor centres and present the results of fitted probability distributions to activity durations in Sect. 3. Next, we develop a simulation model that represents the donation process of a donor centre over a typical day using the randomly drawn realisations from the fitted probability distributions in Sect. 4. In Sect. 5, we explain our simulated annealing algorithm and neighbourhood search method to determine the minimum staffing requirement. In Sect. 6, we discuss the numerical results of our method with an analysis. We conclude the paper by providing a summary of our work and giving some directions for future research in Sect. 7.

2 Literature review

This section addresses the methods of fitting probability distributions to data and the minimum staffing problem using the most relevant literature.

2.1 Data fitting with probability distributions

The histograms generated for the data on the duration of activities registration, questionnaire completion, and assessment are positively skewed (see Fig. 2). Therefore, we need to investigate distributions that are suitable for positively skewed data. Since the distributions

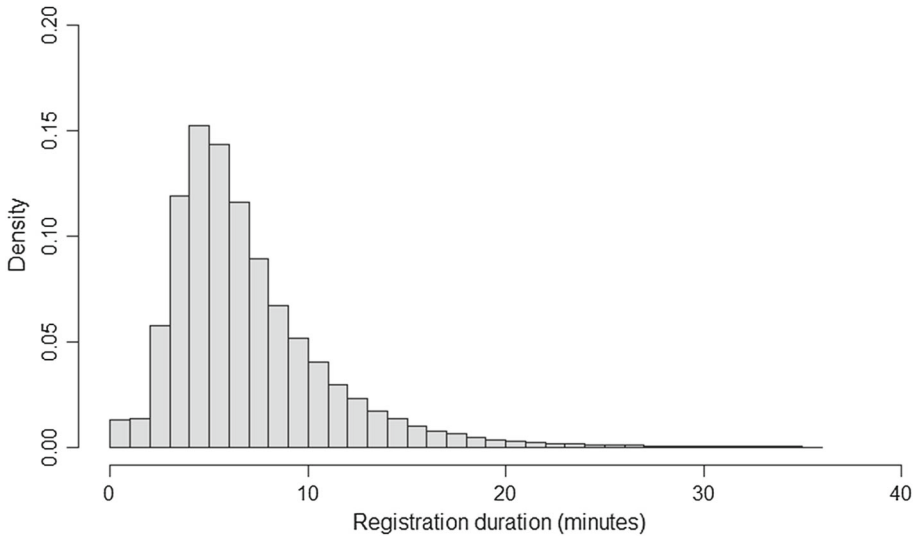


Fig. 2 Histogram of registration duration data of the Melbourne CBD donor centre

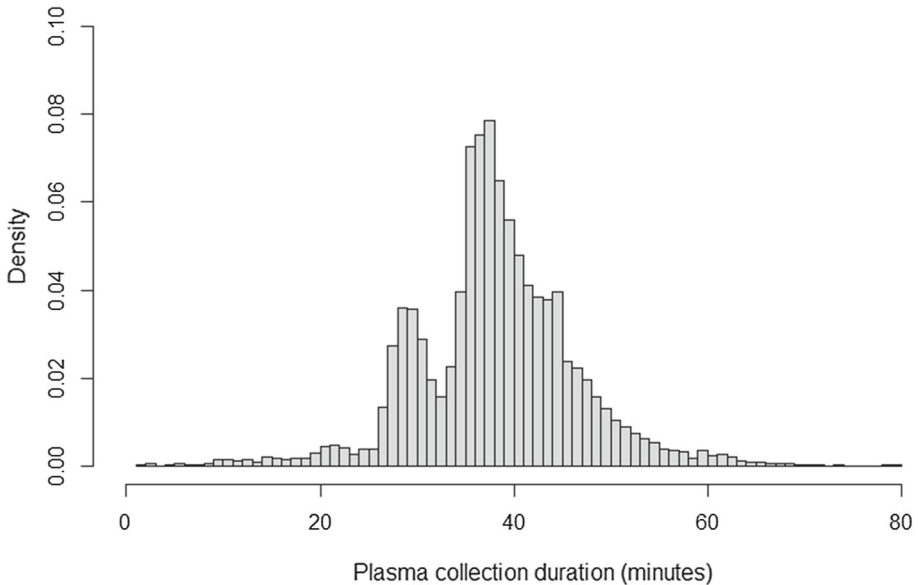


Fig. 3 Histogram of plasma collection duration data of the Gosford donor centre

of collection durations often have two or more peaks (see Fig. 3), we discuss the relevant literature on fitting multi-model distributions to data.

Normal distribution is commonly used for fitting continuous data (Siegel & Wagner, 2022; Mihaylova et al., 2011). When the data is skewed or non-normal, different distributions such as gamma, log-normal, and Weibull are preferred.

Patients' length of stay data is one of the most studied right-skewed distributions in healthcare modelling literature. In particular, gamma, log-normal, and Weibull distributions

were used by McClean and Millard (1993); Marazzi et al. (1998); Gardiner et al. (2014) while Williford et al. (2020) used a gamma mixture distribution for dealing with highly skewed hospital length of stay distributions. Furthermore, Faddy et al. (2009) showed that the phase-type distribution is more suitable for length of stay data after comparison with gamma and log-normal distributions. Kumar and Anjomshoa (2019) also used a Coxian phase-type distribution for length of stay data which is positively skewed.

Lovelace-Tozer (2018) applied gamma, log-normal, Weibull, and Coxian-phase type distributions to find the best fit for data on the duration of all activities at a blood donor centre. The Coxian phase-type distributions were best suited to fit donor centre activity durations, except for plasma, and single-dose and double-dose platelet collections. For plasma and double-dose platelet collection durations, a Gaussian mixture distribution with three components and a gamma distribution were fitted. She used the fitted distributions to develop a simulation model to represent the donation process at a donor centre. However, no distribution was well fitted for single-dose platelet collection duration so she used the empirical density function and bootstrapping observations when developing the model. Phase-type distributions have recently been popular for stochastic modelling of data from many areas, including the healthcare industry, survival analysis, and queueing theory (Fackrell, 2009). In particular, it appears that the application of Coxian phase-type models, which Cox (1955) first proposed, is of great relevance.

The phase-type (PH) distribution, introduced in 1975 by Neuts (1975) is the distribution of the absorption time in a finite-state continuous time Markov chain. The Coxian phase-type distribution is a special type of Markov model and a subclass of phase-type distribution where the transient states are ordered and transitions are forward owing with absorption possible from any of the states (Kumar, 2019). Latouche and Ramaswami (1993) developed a simple algorithm to model inter-arrival times and service times. According to Donnelly et al. (2018), the most common methods for fitting the Coxian phase-type distribution are maximum likelihood estimation and often using the Expectation-Maximisation (EM) algorithm, which was implemented for phase-type distributions by Asmussen et al. (1996).

2.2 Minimum staffing problem

Most studies have shown that determining the minimum number of staff required to cover the demand for service over a tactical horizon is the first step in developing a shift schedule Ernst et al. (2004); Baker et al. (1979). This step is known as *staffing* or *staff dimensioning*. Defraeye and Van Nieuwenhuysse (2016) provided a state-of-the-art literature review on staff scheduling approaches that account for demand. They decomposed the scheduling process into four steps, such as forecasting demand (based on empirical data), determining staffing requirements, shift scheduling, and rostering. The ultimate goal of determining the staffing demand is to reduce costs by developing an optimal shift schedule (Gunawan & Lau, 2013; El-Rifai et al., 2015). In this sub-section, we study the methods developed in the literature for solving the minimum staffing problem.

There are a limited number of studies conducted to determine the minimum staffing requirement at blood donor centres. Van Brummelen et al. (2018) modelled a blood collection site in the Netherlands as a tandem queue with three key stages to determine the average waiting time. They considered whole blood donors who arrived without an appointment. The minimum number of staff members required for any given 30-minute duration was determined using queueing theory.

Our focus is to take into account the appointment-based arrivals with earliness and tardiness, random arrivals, and no-shows, which makes the arrival process complex. Many studies applied queueing theory to determine waiting time by incorporating only the unscheduled arrivals. Hence, there is a lack of literature on modelling complex arrival processes and service processes using queueing theory. A common way of modelling a complex system is to develop simulation models. Most importantly, a simulation model can be used to investigate novel ideas that are challenging to test empirically.

Alfonso et al. (2015) incorporated a simulation-optimisation approach for capacity planning and appointment scheduling of blood donors. However, they also took into account only the unscheduled donor arrivals and no-shows of scheduled donors. Even though their focus is appointment scheduling, this approach gives us an insight into the problem formulation. Smith and Nelson (2015) developed a simulation model to obtain passenger waiting times for flight check-ins at an airport. Hence, there is no study that considers scheduled donors, unscheduled donors, no-shows, and the earliness and tardiness of scheduled donors simultaneously.

Rohleder et al. (2011) used a discrete event simulation model to fine-tune the number of required human resources to match the demand at an outpatient orthopedic clinic. They found that adding some key resources, like an X-ray technician, could reduce waiting times. Wang et al. (2012) used Markov chain modelling to improve the workflow of a tomography department and showed that improving productivity at the bottleneck leads to the highest productivity increase via a what-if analysis. Blake and Shimla (2014) determined the minimum staffing requirements of a blood donor centre using a deterministic flow shop model. These studies show improvements in their systems by adjusting staffing levels appropriately. But they have not optimised the number of staff, which is an important factor for the financial aspect of an organisation.

Ahmed and Alkhamis (2009) presented a discrete event simulation-based optimisation model to determine the optimal number of doctors, lab technicians, and nurses in an emergency department while minimising patient waiting times. Their objective function was to maximise the *throughput* (patients dismissed per unit time) subject to a budget constraint, a constraint imposed on the average waiting time in the system for patients, and upper and lower bounds of the number of servers. The authors used a feasibility detection procedure with stochastic constraints.

The methods of finding the minimum sufficient staff are specific to each problem, as they have different objectives and constraints. Rodriguez et al. (2015) found the minimum number of resources required for each profession to meet all demands in home-care services. They developed two optimisation problems and solved them using branch-and-cut algorithms. The first problem finds the minimum required staff for each resource and each scenario. It has a single objective that minimises the number of routes assigned to resources, along with several constraints for resource allocation. The second problem's objective is to find the minimum sufficient staff to meet the given performance level.

Our study is interested in discrete simulation-based optimisation methods where the objective function must be estimated through stochastic simulation. This is quite an underdeveloped research area. Stochastic optimisation problems can be solved using exact methods such as branch-and-bound, ranking and selection, multiple comparison methods, and also heuristic methods such as evolutionary algorithms, tabu search, ant systems, random search, simulated annealing, and genetic algorithms (Pflug, 2009). Amaran et al. (2016) provided a better review of algorithms for simulation optimisation. Our problem has a large number of potential solutions. Therefore, it is better to choose algorithms such as tabu search, the simulated

annealing algorithm, and the genetic algorithm that have a search component. However, each search strategy has its own set of limitations.

A serious drawback of genetic algorithms is their inefficiency when implemented on sequential machines; hence, they must be implemented on parallel machines (Jog et al., 1991). According to comparison analyses conducted by Connor and Shea (2000) and Kamboj and Sengupta (2009), simulated annealing finds higher quality solutions than tabu search at the cost of increased computational expense. Henderson et al. (2003) explained that while the simulated annealing algorithm has proven to be a good technique for solving difficult discrete optimisation problems, its main disadvantage is that solving a complex system problems can be an extremely slow, albeit convergent, process that consumes significantly more processor time than some conventional algorithms. Therefore, researchers modified the original simulated annealing algorithm of Kirkpatrick et al. (1983) to adapt it to the background of the problem to get more efficient results. Basically, they used two ways to make the simulated annealing algorithm intelligent: annealing schedules to control the temperature (Ahmed 2007; Shi-hua et al. 2016), and solution search strategies (Matsuo et al., 1989; Zhou et al., 2015; Bayram and Sahin, 2013). Therefore, we adapt the simulated annealing algorithm to our minimum staffing problem and develop a novel neighbourhood generating mechanism to improve its efficiency.

The simulated annealing algorithm is popular for its ability to adapt to the context of the problem, its convergence properties, and its ability to escape from local optima via hill-climbing moves. We review recent studies of the simulated annealing algorithm from different contexts since there is no literature that applies the simulated annealing algorithm in the context of a blood donor centre. Suman and Kumar (2006) provided a review of studies on simulated annealing as a tool for single and multi-objective optimisation. Ahmed and Alkhamis (2002) proposed combining the simulated annealing algorithm with the ranking and selection procedure for solving the discrete simulation-based optimisation problem of an inventory system. Bulgak and Sanders (1988) and Jorge and John (1992) integrated simulation with a modified simulated annealing algorithm to find the optimal buffer allocation.

3 Current load of the system

3.1 Data

We received data sets corresponding to the four donor centres of ARCL established in Melbourne CBD, Gosford, Townsville, and Warrnambool. There are two datasets for each donor centre, including donor timestamps and staff rosters. The data sets for the Melbourne CBD donor centre span from 28 February 2017 to 24 December 2018, whereas the data for the other three donor centres span from 28 February 2017 to 13 December 2018.

The timestamp dataset consists of donation identification numbers; a unique number for each donation, including donations from walk-ins. Each row corresponds to a donation with information on donor demographics, blood collection types, timestamps, and the serving nurse. Once all the variables in the timestamp data sets were clarified, we cleaned the data sets by identifying duplicate data rows. There were a few unusual cases where some variables' entries were missing from the datasets. We investigated these unusual cases by discussing them with the ARCL. The reasonable cases, such as donor deferrals due to ineligibility for donations, were retained, and the absurd cases were eliminated.

Table 1 Appointment details of scheduled donors in each donor centre

Donor centre	Melbourne CBD	Gosford	Townsville	Warrnambool
Earliest appointment time	7:00 AM	8:00 AM	7:00 AM	7:00 AM
Last appointment time	7:00 PM	8:00 PM	7:00 PM	7:30 PM

Table 2 Proportions of each blood collection type of scheduled donors

Collection type	Whole blood	Plasma	Single-dose platelets	Double-dose platelets
Melbourne	0.5160	0.4670	0.0040	0.0130
Gosford	0.5090	0.4910	–	–
Townsville	0.0931	0.9069	–	–
Warrnambool	0.5170	0.4830	–	–

The Melbourne CBD donor centre operates seven days a week, while both the Gosford and Townsville donor centres operate from Monday to Saturday. In the Warrnambool donor centre, appointments were scheduled from Monday to Friday. Since the number of appointments and operating times are different between weekdays and weekend days for each donor centre, we consider only the weekday data for the data analysis of this research. Therefore, donors can make their appointments during a weekday from the earliest appointment time to the last appointment time in each donor centre as shown in Table 1.

The Melbourne CBD donor centre collects whole blood, plasma, and single-dose and double-dose platelets, whereas the other three donor centres collect only whole blood and plasma. The proportions of different blood collection types that were analysed from the timestamp data sets in each donor centre are shown in Table 2. The Melbourne CBD donor centre is classified as a metropolitan static donor centre while the others are classified as non-metropolitan static donor centres. The Melbourne CBD donor centre is the largest donor centre in Victoria. The Warrnambool donor centre is a small donor centre that is also established in Victoria. Out of the four donor centres, it has the fewest daily donor appointments, and 51.7% of collections were whole blood, as shown in Table 2. The Gosford donor centre is a medium-sized donor centre situated in West Gosford, New South Wales. The Townsville donor centre is Australia's first plasma-focused donor centre and was established in Queensland. According to the data analysis of the Townsville donor centre as shown in Table 2, only 9.3% of collections were whole blood, and the rest of the collections were plasma. However, starting from December 2017, its whole blood collection was continued only for treating *therapeutic venesection patients*.

3.2 Fitting probability distributions

This Section explains how we identified the most suitable distributions to model the durations of

- *Registration* from registration start time to end time,
- *Questionnaire completion* from registration end time to end time of questionnaire completion,
- *Assessment* assessment start time to end time,
- *Collection of different blood components* needle-in time to needle-out time.

Next, we calculated all these activity durations using the timestamps of donors. All distributions were positively skewed with long tails, except collection durations, as we discussed in Sect. 2.1. Therefore, we tested several probability distributions that are defined for non-negative real values and popular for right-skewed data, such as the gamma, log-normal, Weibull, and Coxian phase-type distributions, to find the best fit for the durations of registration, questionnaire completion, and assessment. The suitability of the normal mixture, gamma mixture, Weibull mixture, and Coxian phase-type distributions for the collection durations was then investigated, given that their histograms had multi-model shapes.

We used the R package *fitdistrplus* for fitting gamma, log-normal, and Weibull distributions to the data (Delignette-Muller and Dutang 2015). The *mapfit* package in R was utilised for fitting Coxian phase-type distributions which uses the Expectation-Maximisation (EM) algorithm for parameter estimation (Okamura & Dohi, 2016). For fitting mixture distributions, we used *mixR* R package (Yu 2022).

We selected the best-fit distribution that has the minimum *Akaike Information Criterion* (AIC) (Akaike, 1973) and *Bayesian Information Criterion* (BIC) (Kriegeskorte, 2015) values to avoid overfitting. These information criteria are tools for selecting the most suitable model among a collection of candidate models to fit to the set of data. They are likelihood-based measures of model fit that use a penalty for the number of parameters in the model. Furthermore, we use one-sample *Kolmogorov-Smirnov* (KS) test (Frank, 1951) where the false positive rate is 0.01 to determine if a sample comes from a population with a specified distribution. If AIC and BIC suggest different models, we choose the model with fewer parameters (Kumar, 2019). However, for the Coxian phase-type distributions, we considered only the AIC values since the MAPfit package does not provide BIC values.

According to the KS test results, the gamma, log-normal, and Weibull distributions did not provide statistically significant fits. Coxian phase-type distributions provided better fits for the registration, questionnaire completion, and assessment durations for all four donor centres. Different mixture distributions were fitted for collection durations. We have given a summary of fitted distributions in Table 3 where the number of components of each distribution can be found within brackets next to the distribution name. Readers may refer to “Appendix A” for information on the number of components, AIC, BIC, log-likelihood, and the KS-test’s p -values of all fitted distributions to activity durations in each donor centre.

Figures 4 and 5 depict the fitted 43-phase Coxian distribution to the data on assessment duration and the fitted 5-component gamma mixture distribution to the data on whole blood collection duration of the Melbourne CBD donor centre, respectively.

4 Discrete event simulation model

This section explains the process of developing a discrete event simulation model that represents the operations of a donor centre during a typical weekday. The simulation model allows us to calculate the mean waiting time associated with a certain staff configuration. We use the developed simulation model to determine the average waiting times of each donor centre by changing the parameter values. We explain the donor attributes that are associated with a donor and the methods of generating them in Sect. 4.1 followed by the integrated model in Sect. 4.2.

Table 3 Fitted probability distributions for each activity durations

Activity Duration	Melbourne	Gosford	Townsville	Warmambool
Registration	(37) Coxian	(9) Coxian	(8) Coxian	(7) Coxian
Questionnaire filling	(98) Coxian	(28) Coxian	(19) Coxian	(15) Coxian
Assessment	(43) Coxian	(17) Coxian	(11) Coxian	(16) Coxian
Whole blood collection	(5) Gamma mixture	(4) Gamma mixture	(4) Gamma mixture	(7) Gamma mixture
Plasma collection	(6) Normal mixture	(7) Normal mixture	(5) Normal mixture	(3) Normal mixture
Single-dose platelet collection	(5) Gamma mixture	–	–	–
Double-dose platelet collection	(5) Normal mixture	–	–	–

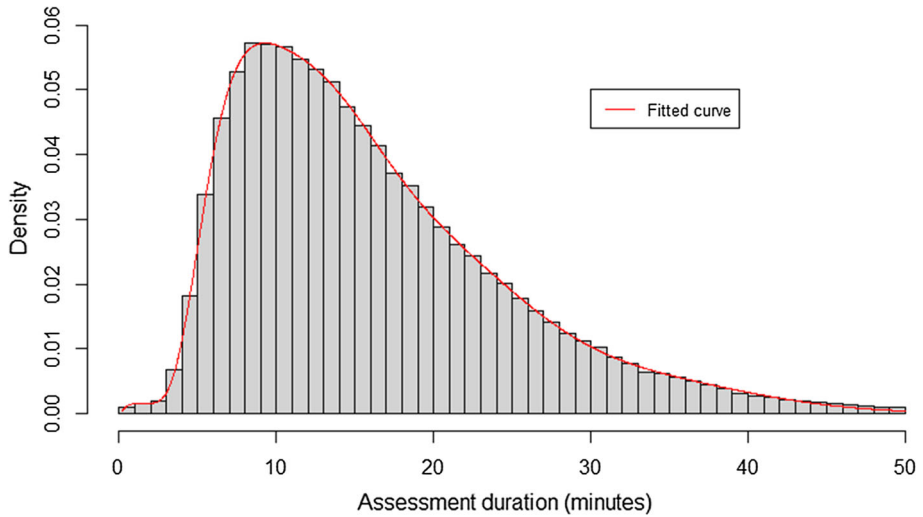


Fig. 4 Histogram of the assessment duration of the Melbourne CBD donor centre with the fitted curve

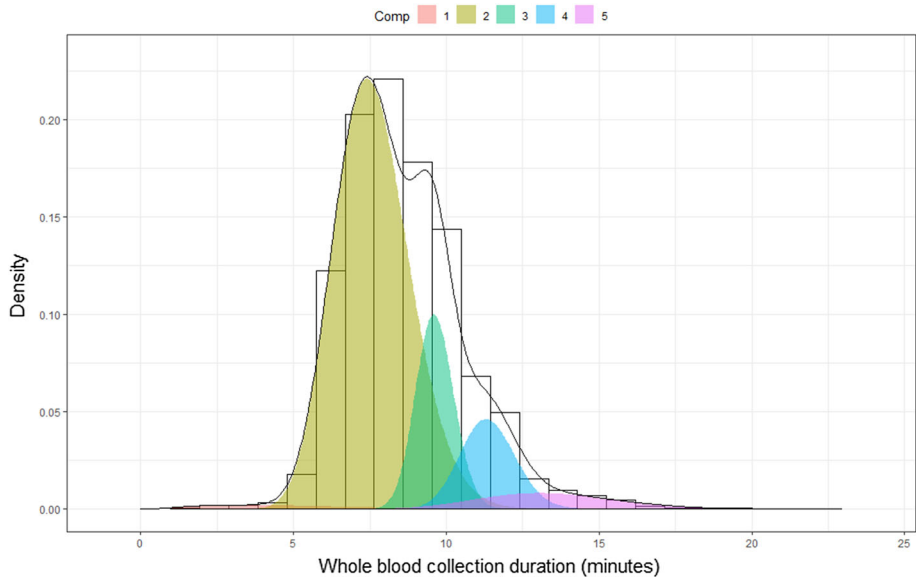


Fig. 5 Histogram of the whole blood collection duration of the Melbourne CBD donor centre with the fitted curve

4.1 Donor attributes

We know the donor centres encourage donors to make appointments before the donation day but accept unscheduled donors if time permits; otherwise, schedule a suitable appointment time on an upcoming day for them. Each donor who arrives at the donor centre with an appointment has the following attributes:

- Appointment time
- Arrival time
- Collection type: whole blood, plasma, single-dose platelets, double-dose platelets
- Registration duration
- Questionnaire completion duration
- Assessment duration
- Needle-in duration
- Collection duration
- Needle-out duration
- Resting duration on the couch
- Resting duration at the refreshment area

Unscheduled donors are those that visit the donor centre without making an appointment; they have all of the attributes listed above, with the exception of the appointment time. Additionally, donors who miss their appointment only have the appointment time and collection type attributes. We must first establish the parameters of each of these donor attributes because they are all stochastic input variables for the simulation model. We now specify the parameters of donor attributes for all four donor centres along with the methods of generating them.

4.1.1 Scheduled donor appointment times

Table 1 shows the parameters required for generating donor appointment times. Each donor centre's appointment times are every 15-minute mark from the first appointment time to the last appointment time. For example, appointment times in the Melbourne CBD donor centre are 7:00 AM, 7:15 AM, 7:30 AM, ..., 7:00 PM. For each donor centre, the number of appointments for a given 15-minute mark was derived by calculating the average number of appointments made at that 15-minute mark. Hence, we scheduled 294, 84, 94, and 53 total donors for the Melbourne CBD, Gosford, Townsville, and Warrnambool donor centres, respectively, for a typical weekday.

4.1.2 Blood collection types for scheduled donors

For each scheduled donor, we generated a blood collection type according to the proportions of each blood collection type of scheduled donors as shown in Table 2. The proportions were calculated using timestamp data. For example, consider the Melbourne CBD donor centre, which has 294 scheduled donors. For each scheduled donor, we generate a random value from a continuous uniform distribution with parameters 0 and 1. If the random value is smaller than the whole blood proportion, the donor becomes a whole blood donor. Otherwise, if the random number is less than the total proportion of whole blood and plasma, then the donor becomes a plasma donor. Following this method, we can allocate a blood collection type for each donor using their cumulative probabilities.

4.1.3 No-show donors

We extracted the details of donors who have an appointment time but no timestamp details from the timestamp data of each donor centre and considered them no-shows. We partitioned the time from the earliest appointment time to the last appointment time into 2-hour periods

Table 4 No-show proportions given the blood collection type and time category for the Melbourne CBD donor centre

Time category	Whole blood	Plasma	Single-dose platelets	Double-dose platelets
7:00 to 9:00	0.1200	0.1885	0.1114	0.1114
9:00 to 11:00	0.1252	0.1622	0.0917	0.0917
11:00 to 13:00	0.1300	0.1583	0.0995	0.0995
13:00 to 15:00	0.1337	0.1668	0.1359	0.1359
15:00 to 17:00	0.1335	0.1391	0.0643	0.0643
17:00 to 19:00	0.1268	0.1539	0.1121	0.1121

Table 5 No-show proportions given the blood collection type and time category for the Gosford donor centre

Time category	Whole blood	Plasma
8:00 to 10:00	0.1451	0.1039
10:00 to 12:00	0.1433	0.0970
12:00 to 14:00	0.1225	0.1035
14:00 to 16:00	0.1459	0.1027
16:00 to 18:00	0.1677	0.1181
18:00 to 20:00	0.1654	0.1350

Table 6 No-show proportions given the blood collection type and time category for the Townsville donor centre

Time category	Whole blood	Plasma
7:00 to 9:00	0.1870	0.0600
9:00 to 11:00	0.1780	0.0570
11:00 to 13:00	0.2000	0.0650
13:00 to 15:00	0.2260	0.0470
15:00 to 17:00	0.2880	0.0690
17:00 to 19:00	0.1660	0.0690

for the Melbourne CBD, Gosford, and Townsville donor centres, and for the Warrnambool donor centre, we partitioned it into 2.5-hour periods (because it schedules donors for 12.5 h). Then, we calculated no-show proportions given the blood collection type and time category as shown in Tables 4, 5, 6, and 7 separately for each donor centre. For each scheduled donor, we assigned the label no-show or show-up according to the no-show proportions using the same random approach as allocating the collection types. Donors with no-show label were eliminated from the list of scheduled donors. Considering the Townsville donor centre, we can see high no-show proportions of whole blood donors compared to plasma donors since there were a small number of whole blood donors scheduled.

4.1.4 Scheduled donor arrival times

Once the no-show donors have been eliminated from the list of scheduled donors, we computed the arrival times of the remaining donors. We calculated donors' earliness and tardiness times from the timestamp data sets by subtracting the appointment time from the arrival time. Therefore, earliness times become negative values and tardiness times become positive val-

Table 7 No-show proportions given the blood collection type and time category for the Warrnambool donor centre

Time category	Whole blood	Plasma
7:00 to 9:30	0.1710	0.0910
9:30 to 12:00	0.1260	0.0850
12:00 to 14:30	0.1710	0.0930
14:30 to 17:00	0.1230	0.0830
17:00 to 19:30	0.1690	0.0940

Table 8 Parameters for generating scheduled donor arrival times

Donor centre	Melbourne CBD	Gosford	Townsville	Warrnambool
Mean (minutes)	3.8037	2.4833	0.2493	3.4585
Standard deviation (minutes)	10.9755	9.9558	11.6739	12.6660

ues. We fitted a normal distribution for each donor centre's earliness and tardiness times. The means and standard deviations of the normal distributions are included in Table 8. For each scheduled donor, we added the realisation generated from the fitted normal distribution to his appointment time to get the arrival time. If a donor's arrival time is later than the last appointment time in Table 1, we replace it with the last appointment time. A donor's arrival time is replaced by the earliest appointment time in Table 1, if it is earlier than the earliest appointment time.

4.1.5 Unscheduled donor arrival times

We generated unscheduled donor arrival times using an inhomogeneous Poisson process. The probability density function $p(t)$ of unscheduled donor arrivals at time t was derived using Kernel Smoothing, where the kernel is the standard normal distribution (Weglarczyk, 2018). Let u_1, u_2, \dots, u_N be the unscheduled arrival times from the data, where N is the total number of data points and v is the bandwidth for kernel density estimation (a real positive number that defines smoothness of the density plot). Therefore,

$$p(t) = \frac{1}{Nv} \sum_{i=0}^N \frac{1}{\sqrt{2\pi}} e^{-(t-u_i)^2/2v^2}.$$

The intensity function $\lambda(t)$ was generated using the probability density function $p(t)$ to determine the instantaneous arrival rate at time t for a time-varying Poisson process (Ross, 2013). Let λ be the average number of unscheduled arrivals per day. Therefore,

$$\lambda(t) = \lambda p(t).$$

To generate the arrival times for an inhomogeneous Poisson process, Lewis' thinning algorithm is applied (Lewis & Shedler, 1979). Table 9 shows the required parameter values for each donor centre. In addition, each donor centre only accepts unscheduled donors between the earliest and last appointment times listed in Table 1.

Table 9 Parameters of generating unscheduled donor arrival times for each donor centre

Parameter	N	ν	λ
Melbourne CBD	3438	1933.23	6
Gosford	640	2480.98	2
Townsville	1485	1686.74	4
Warrnambool	798	1839.22	2

Table 10 Proportions of each blood collection type of unscheduled donors

Collection type	Whole blood	Plasma	Single-dose platelets	Double-dose platelets
Melbourne	0.5136	0.4660	0.0034	0.0170
Gosford	0.6360	0.364	–	–
Townsville	0.1610	0.839	–	–
Warrnambool	0.7260	0.274	–	–

Table 11 Parameters of the normal distributions for needle-in and needle-out times in minutes

	Needle-in			Needle-out		
	Whole blood	Plasma	Platelets	Whole blood	Plasma	Platelets
Mean	5	10	15	5	5	10
Standard deviation	0.5	1	1.5	0.5	0.5	1

4.1.6 Blood collection types for unscheduled donors

For each donor centre, we randomly generated blood collection types of unscheduled donors according to the proportions of each blood collection type of unscheduled donors as shown in Table 10.

4.1.7 Needle-in and needle-out times

The needle-in and needle-out times of all four donor centres were randomly generated from a normal distribution, where the parameters mean and standard deviation were suggested by ARCL with respect to each collection type as shown in Table 11 since data is unavailable.

4.1.8 Resting times on the couch and resting times at the refreshment area

The resting times on the couch and at the refreshment area were randomly generated from normal distributions. We decided the value of the parameter mean according to the information given by (Lovelace-Tozer, 2018). The standard deviations were assumed as shown in Table 12 since data is unavailable.

4.1.9 Activity durations

The durations for registration, completing the questionnaire, assessment, whole blood collection, plasma collection, and single-dose and double-dose platelet collection can be determined by randomly generated realisations from their fitted distributions in Table 3.

Table 12 Parameters of the normal distributions for resting times on couch and at refreshment area in minutes

	Resting time on the couch	Resting time at refreshment area
Mean	10	15
Standard deviation	1	2

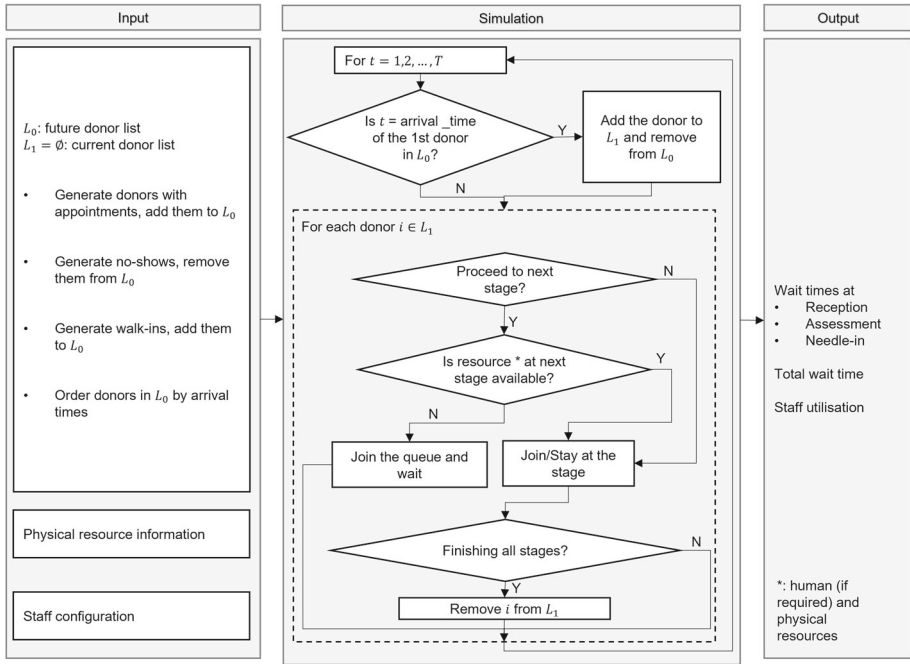


Fig. 6 Diagram of the simulation model

4.2 The integrated model

We developed a comprehensive discrete event simulation model for each donor centre that represents the entire functioning of a donor centre over a typical weekday. Once we generated the donor attributes for each donor, the simulation model was developed as follows. Donors start their donation process according to their arrival time. Next, a queue is implemented for donors to proceed with their registration only when a receptionist is available. Donors start the questionnaire at the end of the registration. Then, donors wait in the queue for assessment until a staff member becomes available. At the blood collection stage, the simulation model checks the availability of a nurse, a couch, and a collection machine for starting the needle-in and then starts the blood collection. When the machine stops the blood collection, the donor must wait for needle-out if all the staff members are busy with the other donors. Except for the needle-in queue, which gives priority to the first donor in the line where the next available machine fits his collection type, all queues follow the first-in-first-out policy. Finally, donors spend their resting time on the couch and at the refreshment area sequentially. Figure 6 illustrates the simulation model.

Table 13 Available physical resources of each donor centre

Physical resource	Melbourne	Gosford	Townsville	Warrnambool
# Registration desks	3	2	2	2
# Interview rooms	10	4	2	3
# Couches	28	9	12	5
# Whole blood machines	7	6	2	5
# Plasma machines	21	6	13	5
# Platelets machines	6	–	–	–

Source: Australian Red Cross Lifeblood

According to Table 13, each donor centre has two to three registration desks, while the number of interview rooms varies between two and four. The Townsville donor centre has two whole blood collection machines since only 9.3% of whole blood donors were scheduled per day. Furthermore, there are 13 plasma collection machines and 12 couches at the Townsville donor centre, which is a higher number considering the number of scheduled appointments due to the fact that plasma collection takes longer than whole blood collection and also because there are many more scheduled plasma donors.

5 Minimum staffing problem

We specify the minimum staffing problem in Sect. 5.1 and our method of determining the minimum staffing requirement during a typical weekday of the Melbourne CBD donor centre in Sect. 5.2. We will test the applicability of the same model to the other three donor centres in Sect. 6.3.

5.1 Problem statement and mathematical model

The minimum staffing problem considered in this study is to determine the number of staff required at each key stage of blood donation over a weekday such that the average waiting time of donors does not exceed some pre-defined target, denoted by w_{target} . The objective of the problem is to minimise the total number of staff hours required. We partition the time horizon of a day into H half-hour intervals, $h \in \{1, 2, \dots, H\}$, and account for three key stages, $j \in \{1, 2, 3\}$, registration, assessment, and blood collection, which require the attendance of staff. We note that the time horizon should be longer than the opening time of the donor centre since donors may be served at various stages after the centre closes. We let $\mathbf{x} = \{x_{j,h}, j \in \{1, 2, 3\}, h \in \{1, 2, \dots, H\}\}$, where $x_{j,h}$ is the number of staff at stage j in half-hour slot h , to represent a staff configuration. Furthermore, we let $w_{avg}(\mathbf{x})$ denote the average donor waiting time for a given configuration \mathbf{x} . The mathematical model of the problem can be formulated as

$$\min_{\mathbf{x}} f_1(\mathbf{x}) = \sum_{j=1}^3 \sum_{h=1}^H x_{j,h} \quad (1)$$

$$\text{s.t. } w_{avg}(\mathbf{x}) \leq w_{target} \quad (2)$$

$$x_{\min,j} \leq x_{j,h} \leq x_{\max,j} \quad \forall h \in \{1, 2, \dots, H\}. \quad (3)$$

The objective function (1) is the total working hours of staff. Constraint (2) ensures the waiting time target is met. Constraint (3) specifies the range of the decision variable $x_{j,h}$. The lower bound $x_{min,j}$ is generally 1 during the opening hours of the center, and can be set to 0 after the opening hours. On the other hand, the upper bound $x_{max,j}$ can be determined by the physical resource limit; e.g., if there are 7 interview rooms, then the maximum number of staff at the interview stage can be set to 7.

Given the donation process is complex and so are the operations in the donor center, it is difficult to derive an analytical form of $w_{avg}(\mathbf{x})$. Therefore, we apply the Monte Carlo simulation developed in the previous section to generate a large number (D) of scenarios, each with a probability of $1/D$, and approximate the mean of the average waiting time. One simulation corresponds to one scenario or one day of the operation of the donor center. To simplify the constrained model above so that we can develop a search algorithm to solve it, we convert constraint (2) to a penalty term with a penalty coefficient $\alpha \gg 0$ in the objective function, namely, $\alpha(w_{avg}(\mathbf{x}) - w_{target})^+$ with $(z)^+ = z$ if $z > 0$ and $(z)^+ = 0$ otherwise. The final formulation is given as follows.

$$\min_{\mathbf{x}} \sum_{j=1}^3 \sum_{h=1}^H x_{j,h} + \alpha \left(\frac{1}{D} \sum_{i=1}^D w_{avg,i}(\mathbf{x}) - w_{target} \right)^+ \tag{4}$$

where $w_{avg,i}(\mathbf{x})$ is the average donor waiting time in the i th scenario (simulation) for a given \mathbf{x} . The range constraint (3) can be realised by appropriately setting the search space of \mathbf{x} of the search algorithm. The penalty term will guide the search and must be zero for the final solution found by the search algorithm in order to guarantee feasibility.

5.2 Methodology

We propose an adapted simulated annealing algorithm integrated with the Monte Carlo simulation, developed in Sect. 4, to solve the minimum staffing problem discussed above. The inputs of the adapted simulated annealing algorithm are the maximum temperature T_{max} , minimum temperature T_{min} , annealing constant a and the number of simulation replications parameter D .

Figure 7 shows the flowchart of the simulation-based simulated annealing algorithm (SBSA). The inputs of the SBSA include \mathbf{x}_{new} , $\mathbf{x}_{current}$, and \mathbf{x}^* be the new, current, and current best staff configurations (solutions), respectively. As shown in Fig. 7, this flowchart has two loops. The inner loop includes the donor flow simulation with respect to a given staff configuration \mathbf{x}_{new} , and it yields the average waiting time $w_{avg}(\mathbf{x}_{new})$ and further the value of the objective function (4) after repeating the simulation D times. The outer loop represents the simulated annealing algorithm. Initially, the algorithm generates the new staff configuration \mathbf{x}_{new} randomly, restricted to the available physical resources in each stage, such that $x_{min,j} \leq x_{j,h} \leq x_{max,j}$. Furthermore, the objective function values of the current staff configuration $\mathbf{x}_{current}$ and current best staff configuration \mathbf{x}^* are set to be infinity (100,000 is chosen as “infinity”). This ensures the initial \mathbf{x}_{new} becomes both the current and the current best staff configurations during the first iteration of the outer loop. Furthermore, the current temperature T is initially set as the maximum temperature T_{max} .

The outer loop compares the objective function values of the new and current configurations. If $f(x_{new})$ is lower, then it is accepted. Otherwise, it is accepted with a probability p determined by

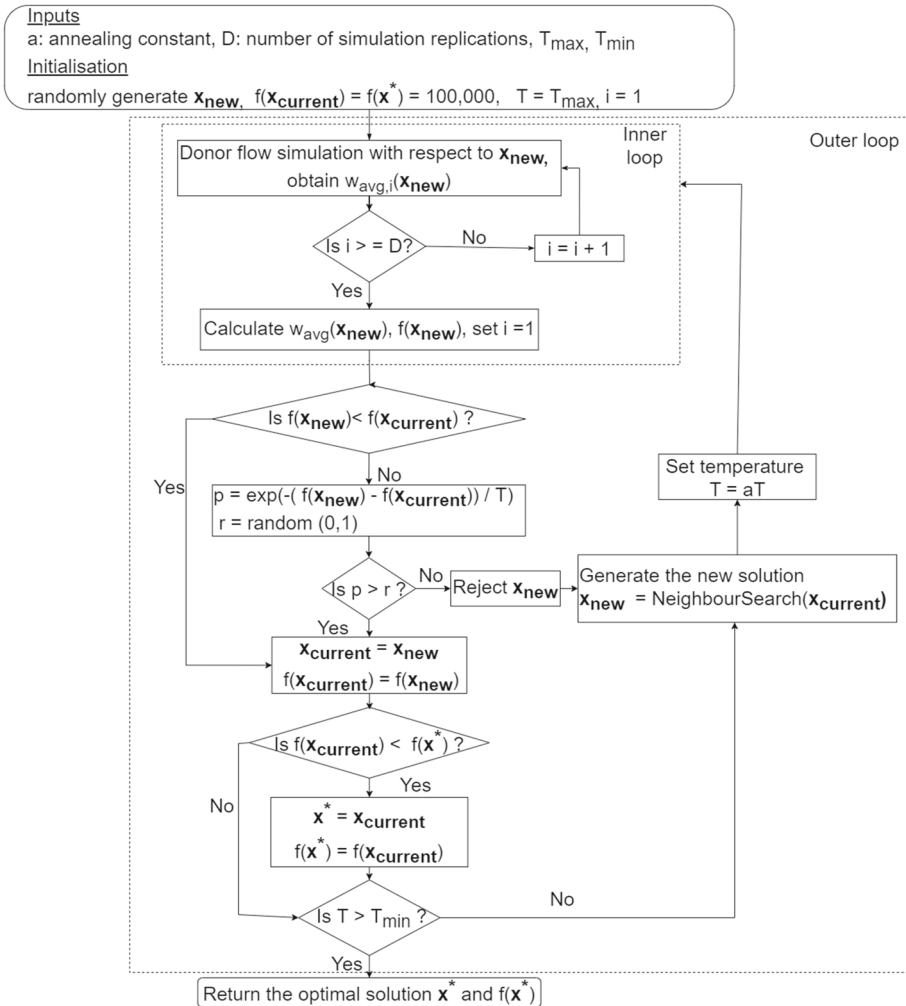


Fig. 7 Flowchart of the simulation-based simulated annealing algorithm

$$p = \exp\left(\frac{-(f(x_{new}) - f(x_{current}))}{T}\right) \quad (5)$$

The accepted configuration then becomes the current configuration $x_{current}$, and the current best configuration x^* is updated if the current solution is better. Next, the simulated annealing algorithm generates a new configuration based on the updated current configuration (or the same current configuration if x_{new} is rejected), utilising the *NeighbourSearch* (NS) algorithm, which is explained in the next sub-section. These steps are repeated until the current temperature T reaches its minimum T_{min} , a parameter close to zero. In each iteration, T is multiplied by the cooling constant a , a parameter close to one,² to generate the temperature for the next iteration. Most importantly, the initial temperature T_{max} should be large enough to explore a large number of configurations.

² Typically in the range from 0.9 to 0.99.

5.3 NeighbourSearch (NS) algorithm

Due to the large solution space of this problem, the simulated annealing algorithm needs to move systematically through it to seek the optimal staff configuration. It requires a neighbourhood search method that generates new staff configurations efficiently and effectively.

The idea behind the proposed NS algorithm is to include both random and informed searches. The informed searches aim to allocate more staff to the stage and/or time slot where (donor) congestion occurs and remove staff where redundancy occurs. To quantify the congestion/redundancy, we define the staff occupancy level as the fraction of time that staff are occupied.

We define binary variable $y_{j,l,h,d}$ and have $y_{j,l,h,d} = 1$ to represent if all the staff allocated to attend stage $j \in \{1, 2, 3\}$ at the $l \in \{1, 2, \dots, 30\}$ th minute of time slot $h \in \{1, 2, \dots, H\}$ of day $d \in \{1, 2, \dots, D\}$ are busy. Then, the occupancy status of a time slot is represented by another binary variable $z_{j,h,d}$ defined based on $y_{j,l,h,d}$ as

$$z_{j,h,d} = \begin{cases} 1, & \text{if } \frac{\sum_{l=1}^{30} y_{j,l,h,d}}{30} > \gamma \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where γ is a parameter between zero and one. Namely, if more than 80% of minutes ($\gamma = 0.8$) in the h th half-hour of (j, h) slot are fully occupied, the slot is considered as busy (or highly occupied). The sample average, denoted by $o_{j,h}$, is the average over the D scenarios/simulations.

A threshold is required to decide whether $o_{j,h}$ is high or low and then help to adjust $x_{j,h}$ in $\mathbf{x}_{\text{current}}$ accordingly to generate \mathbf{x}_{new} . Let ψ be the parameter *occupancy threshold*. Suppose we have the current solution $\mathbf{x}_{\text{current}} = \{x_{1,1}, x_{1,2}, \dots, x_{j,h}, \dots, x_{3,H}\}$ in our simulated annealing algorithm, and the corresponding average occupancy is $o_{j,h}$. Then the new solution, $\mathbf{x}_{\text{new}} = \{x_{1,1}^{\text{new}}, x_{1,2}^{\text{new}}, \dots, x_{j,h}^{\text{new}}, \dots, x_{3,H}^{\text{new}}\}$, is generated based on the following formula.

$$x_{j,h}^{\text{new}} = \begin{cases} x_{j,h} + 1, & \text{if } o_{j,h} > \psi\% \text{ and } x_{j,h} < x_{\text{max},j} \\ x_{j,h} - 1, & \text{if } o_{j,h} = 0 \text{ and } x_{j,h} > x_{\text{min},j} \\ x_{j,h}, & \text{otherwise} \end{cases} \tag{7}$$

When $o_{j,h}$ is higher than the occupancy threshold ψ , the new staff configuration will have one extra staff member in the (j, h) slot compared to the current solution. If the average occupancy is low, $o_{j,h} = 0$, the new solution will have one member fewer in the (j,h) slot. Otherwise, the staff number allocated to the slot remains unchanged. We remark that $o_{j,h} = 0$ implies $z_{j,h,d} = 0 \forall d \in D$. If for at least one day, the slot is busy, $\sum_{l=1}^{30} y > 0.8 \times 30$ and hence $z_{j,h,d}$, then reducing the staff allocation will not occur. When updating staff, we need to consider the staff limitations as well.

Algorithm 1 presents the NS algorithm, which generates a new staff configuration \mathbf{x}_{new} based on the current solution $\mathbf{x}_{\text{current}}$. Initially, it updates the new configuration by assigning the current configuration and then calculates $o_{j,h}$ with respect to $\mathbf{x}_{\text{current}}$. Next, it performs three operations as follows. The first operation (from line 3-20) is the informed search, which adjusts the solution based on Equation (7). The other two operations are random searches, and one of them is performed at a time. The second (from line 19-27) swaps staff numbers from two (j, h) slots, and the third (from line 25-33) operation randomly selects a stage j'

Algorithm 1 NeighbourSearch**Input:** $x_{\text{current}}, x_{\text{max},j}, x_{\text{min},j}$ **Output:** x_{new} 1: Initialization: $x_{\text{new}} \leftarrow x_{\text{current}}$ 2: calculate $o_{j,h}$ for x_{current}

▷ Operation 1: adjusting staff allocation according to Equation (7)

3: **for all** element $j \in \{1, 2, 3\}$ **do**4: randomly generate an integer n in the interval $[0, H]$ 5: **if** $n > 0$ **then**6: randomly generate a list “index” of n unique integers from the interval $[1, H]$ 7: **for all** element $h \in \text{index}$ **do**8: **if** $o_{j,h} > \psi\%$ and $x_{j,h}^{\text{new}} < x_{\text{max},j}$ **then**9: $x_{j,h}^{\text{new}} \leftarrow x_{j,h}^{\text{new}} + 1$ 10: **else**11: **if** $o_{j,h} = 0$ and $x_{j,h}^{\text{new}} > x_{\text{min},j}$ **then**12: $x_{j,h}^{\text{new}} \leftarrow x_{j,h}^{\text{new}} - 1$ 13: **end if**14: **end if**15: **end for**16: **end if**17: **end for**18: randomly generate an integer “swapping” in the interval $[0, 1]$ 19: **if** $\text{swapping} = 1$ **then**

▷ Operation 2: swapping two staff numbers

20: randomly choose a stage j_1 21: randomly choose two different time slots h_1 and h_2 22: $\text{temp} \leftarrow x_{j_1, h_2}^{\text{new}}$ 23: $x_{j_1, h_2}^{\text{new}} \leftarrow x_{j_1, h_1}^{\text{new}}$ 24: $x_{j_1, h_1}^{\text{new}} \leftarrow \text{temp}$ 25: **else**

▷ Operation 3: replacing a staff number

26: randomly choose stage j' 27: randomly choose time slot h' 28: randomly generate an integer β from the interval $[\min(x_{j',1}^{\text{new}}, x_{j',2}^{\text{new}}, \dots, x_{j',H}^{\text{new}}), \max(x_{j',1}^{\text{new}}, x_{j',2}^{\text{new}}, \dots, x_{j',H}^{\text{new}})]$ 29: $x_{j',h'}^{\text{new}} \leftarrow \beta$ 30: **end if**

and a time slot h' and randomly sets the number of staff members for the (j', h') slot subject to the current values of $x_{j',h} \forall h$.

6 Numerical results

Focusing on the Melbourne CBD donor centre, we determine the parameters of the SBSA algorithm in Sect. 6.1 and then obtain the minimum staffing requirements in Sect. 6.2. Next, using the same parameters, we further investigate our SBSA algorithm under different scenarios in Sect. 6.3 and also test the applicability of the model to the other three donor centres in Sect. 6.4.

The model was implemented in Python (3.8.6). Experiments were carried out on SPARTAN (Lafayette et al., 2016) with one CPU. We designed our experiments to provide pragmatic support for our method and to find the minimum staffing requirement of the donor centre.

Table 14 Testing for the number of replications of the simulation

D	Average computational time in hours	Average relative error
25	15.87	6.5%
50	28.01	1.5%
75	41.15	1.4%
100	53.76	–

6.1 Parameter selection for the SBSA algorithm

In this section, we determine the parameters of the SBSA algorithm, focusing on the Melbourne CBD donor centre. We fine-tune key parameters, including the sample size D , occupancy threshold ϕ , maximum temperature T_{max} , annealing constant a while keeping parameters $T_{min} = 0.1$, $\gamma = 0.8$ and $\alpha = 10,000$ fixed so that the simulated annealing algorithm can run at a reasonable time and still produce optimal/near-optimal solutions. The waiting time target w_{target} is 6 min as per our explanation in Sect. 1.

Initially, we set the total number of half-hours $H = 34$ (so the operating duration is 17 h). We will then identify the actual operating duration by finding the minimum staffing requirement. When generating neighbour solutions using Algorithm 1, we set the range of staff numbers $[x_{min,j}, x_{max,j}]$ as [1,3], [1,10], and [1,28] for the registration, assessment, and blood collection stages, respectively. Assuming there must be at least one staff member available in each stage during all half-hours, the lower limit is set to one. The upper limit is set to the number of available reception desks, interview rooms, and couches for the registration, assessment, and blood collection stages, respectively.

We narrow down the range for each stage when generating the initial staff configuration, which reduces the variability of staff numbers and the run time of the simulated annealing algorithm. Therefore, $x_{j,h}$ for all (j, h) slots is randomly generated from the intervals [2,3], [4,9], and [4,8] for the registration, assessment, and blood collection stages, respectively, to determine the initial \mathbf{x}_{new} . Furthermore, we always run our SBSA algorithm for ten trials for our experiments.

6.1.1 Sample size

When using the discrete event simulation for waiting time estimations, simulation error is involved. Increasing the sample size (number of replications) can reduce this error, however it is computationally expensive (Nasr & Taaffe, 2013). To find an appropriate sample size D that balance the simulation accuracy and the computational efficiency, we experimented several cases with different D values in the range [25, 100] and compared them by computing the relative error of the average waiting times as a percentage with respect to $D = 100$. Specifically, we used 100 randomly generated staff configurations where each staff configuration is denoted by \mathbf{x}_i such that $\{i = 1, 2, \dots, 100\}$ and their average waiting times $w_{avg,D}(\mathbf{x}_i)$ were determined by repeating the simulation model for D times. Then, we calculate the average relative error as

$$\text{Average relative error} = \sum_{i=1}^{100} \frac{|w_{avg,D=100}(\mathbf{x}_i) - w_{avg,D}(\mathbf{x}_i)|}{w_{avg,D=100}(\mathbf{x}_i)} \% \quad (8)$$

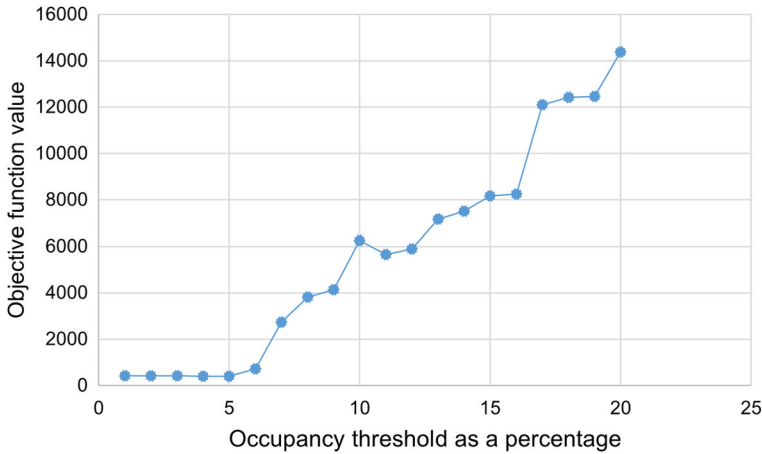


Fig. 8 Occupancy threshold versus objective function value with $T_{max} = 50,000$ and $\psi = 0.8$

Table 15 Annealing schedules

Schedule	1	2	3	4	5	6
a	0.8	0.9	0.8	0.9	0.9	0.9
T_{max}	50,000	50,000	10,000	10,000	5,000	250

In addition, we determined the average computational time in each scenario of D as shown in Table 14. The relative error when $D = 25$ is comparatively higher, even though it saves more than 37 hours of run time. Moreover, $D = 75$ has the smallest relative error, but only marginally better than $D = 50$. However, the computational time of $D = 75$ is significantly longer than $D = 50$. Hence we choose $D = 50$ for our experiments.

6.1.2 Occupancy threshold

For each $\psi = [1, 20]$, we chose the staff configuration that has the minimum objective function value among results of the ten trials. Figure 8 shows the objective function values of those optimal staff configurations with ψ . In the interval $1 \leq \psi \leq 5$, we can see that the objective function values are small and generally insensitive to the choice of ϕ . Additionally, the resulting average waiting times do not exceed the waiting time target. When $\psi > 5$, the objective function values dramatically increase. Therefore, the following experiments were conducted for $1 \leq \psi \leq 5$.

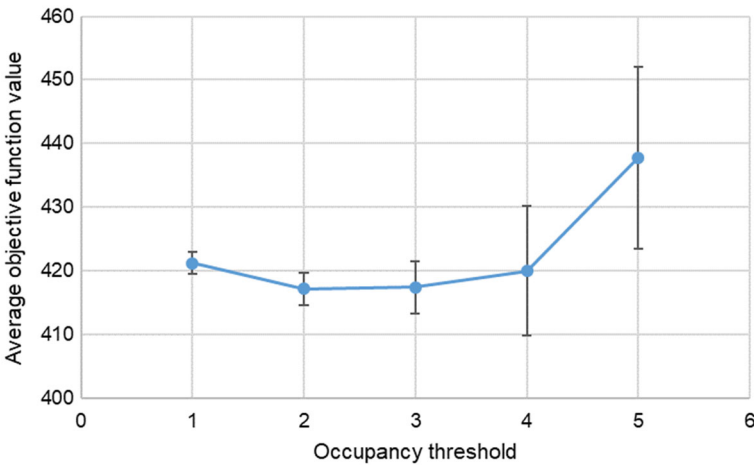
6.1.3 Annealing schedules

Table 15 summarises the parameters for each annealing schedule in terms of the annealing constant a and the maximum temperature T_{max} .

For each annealing schedule, and each occupancy threshold $\psi = [1, 5]$, we executed the simulated annealing algorithm for ten trials and reported the minimum objective function values and the computational times. We calculated the average and standard deviation of computational time for each annealing schedule using the computational times of ten trials.

Table 16 The average and standard deviation of computational times (in hours) of all annealing schedules for each occupancy threshold $\psi = [1, 5]$

T_{max}, a	$\psi = 1$	$\psi = 2$	$\psi = 3$	$\psi = 4$	$\psi = 5$
50,000, 0.8	24.17 \pm 1.43	23.75 \pm 1.33	25.17 \pm 1.76	23.50 \pm 1.93	24.03 \pm 1.61
50,000, 0.9	49.65 \pm 2.26	47.61 \pm 0.46	47.93 \pm 0.83	48.68 \pm 1.99	48.19 \pm 1.22
10,000, 0.8	15.14 \pm 2.31	14.78 \pm 1.74	15.38 \pm 2.77	14.27 \pm 2.69	13.77 \pm 1.94
10,000, 0.9	42.54 \pm 2.80	42.11 \pm 1.55	46.73 \pm 2.12	42.68 \pm 2.58	44.86 \pm 2.00
5,000, 0.9	46.15 \pm 1.69	40.47 \pm 2.00	45.24 \pm 2.03	40.62 \pm 1.75	39.03 \pm 1.89
250, 0.9	29.06 \pm 2.67	26.10 \pm 2.70	27.70 \pm 1.93	30.85 \pm 2.11	24.98 \pm 2.30

**Fig. 9** Occupancy threshold versus average objective function value with $T_{max} = 10,000$ and $a = 0.8$

The minimum objective function values of different annealing schedules are similar, ranging from 410 to 422. Therefore, we look at their computational time performance in order to choose the best annealing schedule. Table 16 present the average and standard deviation of the run times (in hours) for all annealing schedules under each occupancy threshold. The annealing schedules ($T_{max} = 50,000, a = 0.8$) and ($T_{max} = 10,000, a = 0.8$) are computationally less expensive than the rest of the schedules since they run for a small number of iterations (outer loop of the simulated annealing algorithm). As we mentioned above, the objective function values are relatively stable. Hence, we selected ($T_{max} = 10,000, a = 0.8$) as the annealing schedule for our method since it has the lowest computational requirement with low variability over all occupancy thresholds.

So far we have identified a suitable range for ψ , and we next need to choose the best value of ϕ from that range. We calculated the average objective function value from ten trials of the simulated annealing algorithm for each ψ with respect to the annealing schedule ($T_{max} = 10,000, a = 0.8$). As shown in Fig. 9, the lowest average objective function value with a small variability is obtained by $\psi = 2$. Hence, we set $\psi = 2$ for the simulations in the remaining of the section.

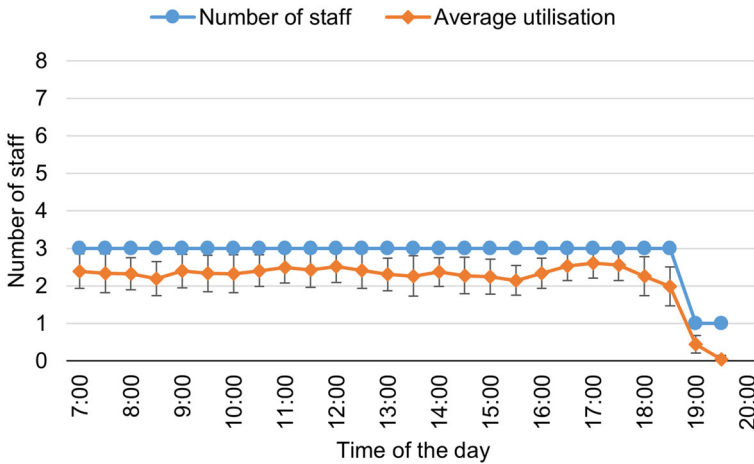


Fig. 10 Minimum staffing requirement with their average utilisation over the day—registration stage

6.2 Minimum staffing requirement

We input the parameters $T_{min} = 0.1$, $\alpha = 10,000$, $T_{max} = 10,000$, $a = 0.8$, $D = 50$ and $\psi = 2$ into our SBSA algorithm to determine the minimum staffing requirement. Furthermore, we always run our SBSA algorithm for ten trials and choose the optimal staff configuration that has the minimum objective function value among them.

Figures 10, 11 and 12 show the number of staff required in each (j, h) slot of the optimum staff configuration. The average number of staff members utilised to serve donors in each slot with respect to the best staff configuration was computed. Figures 10, 11 and 12 show the average staff utilisation with respect to the minimum staffing requirement, while the error bars represent the standard deviations. Although we set $H = 34$ at the beginning, we remove the final few half-hours where their utilisation is zero and staffing demand is only one, which is the lower limit of the number of staff. Hence, the optimal operating duration for each stage is determined as shown in Figs. 10, 11 and 12.

According to Fig. 10, the registration stage requires three receptionists during the first 12 h after opening the donor centre, and only one staff member is enough in the subsequent hour. This result is sensible since there is no donor arrival after 12 h from the beginning. Furthermore, after 13 h, we do not need any staff at reception as all donors have finished their registration by the 13th hour.

The minimum number of staff required at the assessment stage is shown in Fig. 11. During the first half-hour of the day, only three staff members are needed to serve the donors who finish the registration and questionnaire. Next, it requires seven or eight staff members, which is much more than that at the registration stage since the assessment duration is longer than the registration duration. Furthermore, in the middle of the day (from the second half-hour to the 23rd half-hour), the range of average staff is much smaller. Similarly, we can see this feature in the collection stage. In the assessment and collection stages, staff are required for 28 and 30 half-hours, respectively. At the collection stage, the first half-hour requires only one staff member since it takes some time for donors to start blood collection after their arrival, and the second half-hour requires only four staff. Because the needle-in and needle-out times are shorter than the assessment times, only five to six nurses are needed at the blood collection stage during the day, except for the first and last three hours. In all three

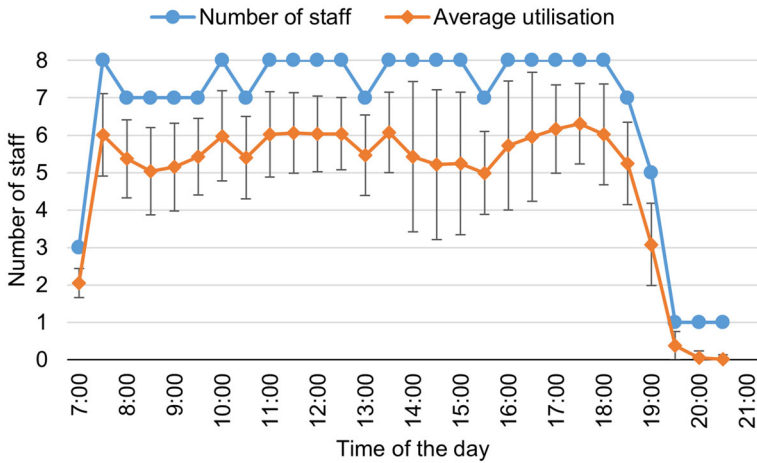


Fig. 11 Minimum staffing requirement with their average utilisation over the day—assessment stage

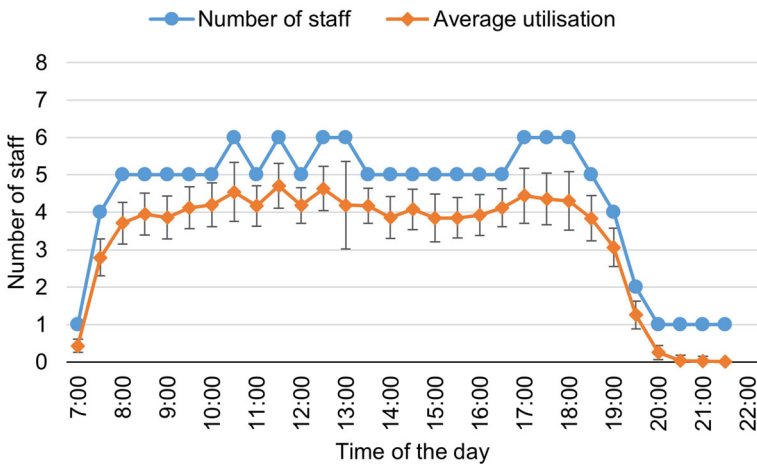


Fig. 12 Minimum staffing requirement with their average utilisation over the day—collection stage

stages, the minimum staffing requirement is reduced to the lower bound of staff during the final hours. Considering the above results, the minimum staffing requirements are sensibly compliant with the donation process.

Almost all half-hours are utilised more than 70% on average in each stage. The final three half-hours of each stage are less utilised, and they are very close to zero. The error bars show the spread of staff utilisation around the average staff utilisation. Specifically, we noticed that the minimum staffing requirement at the registration has small variation in the average number of recipients and also small error bars of utilisation. Therefore, it is important to check for the need for more reception desks at the registration stage, and we will investigate this in the next sub-section.

The total working hours is 196.5 resulting in an average waiting time of 5.99 min, very close to the target. The waiting time distribution is positively skewed, as shown in Fig. 13, and the standard deviation is 7.42 min. Recalling Sect. 3.2, all activity durations are positively skewed with long tails, which leads to a positively skewed waiting time distribution. This simulation-based approach could be easily adapted to account for other key performance

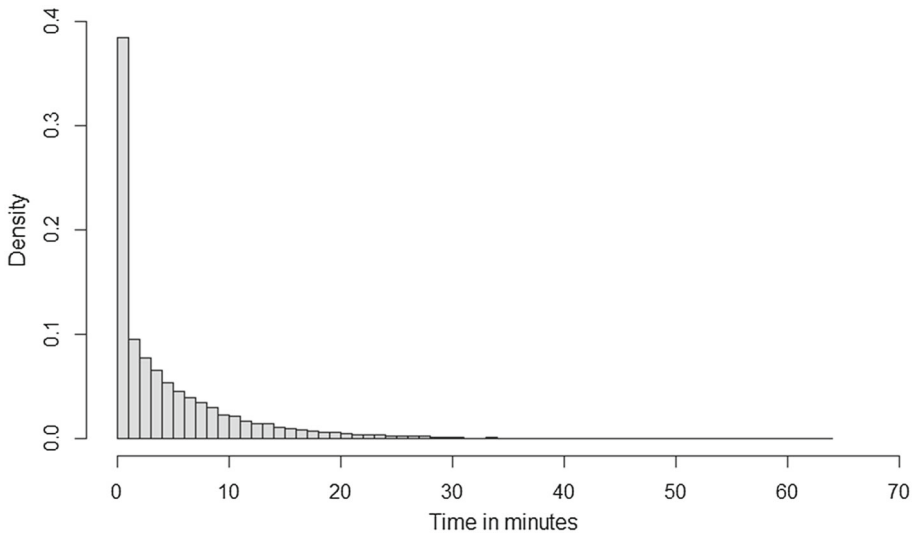


Fig. 13 Waiting time distribution of the Melbourne CBD donor centre under the optimal staffing configuration

indicators. For example, given that the waiting time distribution is right-skewed, the ARCL could consider the median or 90th percentile of the waiting time as a key performance indicator. In that case, one simply needs to calculate the median, or 90th percentile, of the waiting time from the simulations.

6.3 Other applications of the SBSA approach

In this section, we further investigated the applications of our adapted simulated annealing algorithm that is integrated with simulation using new benchmarks. We keep using the parameters $T_{min} = 0.1$, $\alpha = 10,000$, $T_{max} = 10,000$, $a = 0.8$, $\gamma = 0.8$ and $\psi = 2$ fixed. Again, we repeated the simulated annealing algorithm ten times and obtained the rounded average staff and its average waiting time for the following investigations.

The first scenario that we investigate is to assess the impact of the physical resources appropriately to investigate the changes in total staff and the average waiting time. As discussed in the previous subsection, the registration stage is a bottleneck in the donation process due to the limited number of reception desks. So we increased the number of reception desks and compared the resulting average waiting time and minimum staff working hours in Table 17. The waiting time target is still 6 min, and when the number of desks is equal to or greater than 3, this target is achievable. The findings indicate that when there are 4 reception spaces available, compared to the current practice, the average waiting time reduces by 14.4 s (approximately 4%) and the total working hours grow. Moreover, the standard deviation is significantly reduced from 7.42 min to 6.30 min. However, having more reception desks does not significantly improve the waiting time or standard deviation but raises the total working hours. So, if Lifeblood wishes to improve the waiting time and/or reduce the variance in waiting time, increasing the reception spaces is an effective solution.

Furthermore, we discovered that if the maximum staff at the registration stage is reduced to two, it is impossible to keep the average waiting time below the target. We did not investigate the effects of varying the numbers of interview rooms, couches, and collection machines

Table 17 Information on average waiting time, its standard deviation, and total working hours under different numbers of reception spaces

Number of spaces	Total working hours	w_{avg} (min)	Standard deviation (min)
3	196.5	5.99	7.42
4	200.5	5.75	6.30
5	205.5	5.72	6.36
6	207.0	5.69	6.25
7	208.0	5.68	6.02

Table 18 Information of average waiting time, its standard deviation and total working hours under different waiting time targets

w_{target} (min)	Total working hours	w_{avg} (min)	Standard deviation (min)
5	217.0	4.90	5.77
6	196.5	5.99	7.42
7	194.5	6.70	7.16
8	183.0	7.82	8.59
9	176.5	8.80	9.67

Table 19 Total staff, average waiting times and its standard deviations when increasing the registration duration

Percentage of increment	Total working hours	w_{avg} (min)	Standard deviation (min)
5	209.0	6.00	7.01
10	214.5	6.20	7.56

since they are already more than the minimum staffing requirement. That is to say, there exists some redundancy in terms of physical resources at the interview and blood collection stages. In the case of establishing a new donor center or having a major update for an existing one, our model and solution approach can be utilised to find the optimal resource allocation.

The second investigation that we conducted was the impact of different waiting time on the total working hours. As shown in Table 18, as the target increases, the total working hours of the resulting minimum staffing requirements decrease. Having a closer look at the result, to reduce the current waiting time target by 1 min requires an extra 20.5 h (approximately 10.4%), while to loosen the target by 1 min merely saves 2 h (approximately 1%). The average waiting times are always close to their target waiting times. This demonstrates that the annealing schedule ($T_{max} = 10,000$, $a = 0.8$) we chose is effective in other benchmarks as well.

Finally, we tested the effect of increased registration, assessment, and collection durations separately by some percentage to test the effect on total working hours and average waiting time of the minimum staffing requirements. When the registration duration is extended by 10% or more, it is impossible to meet the target waiting time as shown in Table 19. When the percentage of the increment is 5%, the average waiting time is 6 min. This result is align with the previous observation that the reception stage is the bottleneck stage.

Table 20 shows that the minimum staffing requirement is increased rapidly when assessments are prolonged. Furthermore, it is impossible to meet the 6-min target if the average assessment time is incremented by 40% or more.

Table 20 Total working hours, average waiting times and its standard deviations when increasing the assessment duration

Percentage of increment	Total working hours	w_{avg} (min)	Standard deviation (min)
5	202.0	6.00	6.82
10	207.5	5.91	6.76
15	214.0	5.98	6.74
20	220.0	5.97	6.74
30	228.0	5.98	6.78
40	231.5	6.09	7.24

Table 21 Total staff working hours, average waiting times and its standard deviation when increasing the collection duration

Percentage of increment	Total working hours	w_{avg} (min)	Standard deviation (min)
5	205.5	6.00	6.89
10	213.5	6.00	7.03
15	221.0	6.00	7.18
20	226.0	6.12	7.93

According to Table 21, we can extend the collection durations by up to 15% and still meet the target waiting time. The standard deviations increase as the average waiting time increases, as shown in Tables 19, 20 and 21. It seems that the standard deviation is not sensitive to the increment in assessment duration compared to the collection duration.

6.4 Other donor centres

Previously, we tested our simulation-based simulated annealing algorithm using several scenarios for the Melbourne CBD donor centre. In this section, we test the ability of our method to be applied to the Gosford, Townsville, and Warrnambool donor centres.

Similar to the Melbourne CBD donor centre, the operating durations of the Gosford, Townsville, and Warrnambool donor centres are partitioned into half-hour periods, and so we can find the required staff numbers for each half-hour slot of each key stage. We utilised our adapted simulated annealing algorithm with the same parameters determined in Sect. 5. The initial staff configurations are generated at random from the intervals [1,2], [1,2], and [1,3] for the registration, assessment, and blood collection stages, respectively, for all three donor centres. When generating neighbour solutions using Algorithm 1, the bounds for the number of staff members at each stage are set similarly to the Melbourne case study.

For the optimal solutions found by the SBSA algorithm, the average waiting times at the Gosford, Townsville, and Warrnambool donor centres are 5.95 ± 7.12 , 5.98 ± 8.00 , and 5.98 ± 7.79 min, respectively. The standard deviations are greater than the average due to the skewness of the waiting time distribution, as we discussed in Sect. 6.2. Compared to the Melbourne CBD donor center, the waiting time deviations of the other three centres are almost similar.

The summary of the minimum staffing requirements is given in Table 22. Similar to the results for the Melbourne center, it indicates that employees are allocated a few extra hours following the last appointment time of each donor centre. For example, the registration stage of the Gosford donor centre requires staff for 13 h. Considering that registration durations are

Table 22 Summary of the minimum staffing requirements for all four donor centres

Donor centre	Key stage	Total working hours
Melbourne	Registration	37.0
	Assessment	93.5
	Collection	66.0
Gosford	Registration	20.0
	Assessment	29.0
	Collection	24.5
Townsville	Registration	22.5
	Assessment	25.5
	Collection	34.5
Warrnambool	Registration	15.0
	Assessment	22.0
	Collection	20.0

produced from positively skewed probability distributions, they can be longer than 30 min. Therefore, it must be attended for an additional hour beyond the final appointment time.

According to Table 22, all the centres except Townsville dedicated more staff hours to the assessment stage than the other stages because the assessment takes longer than registration and the total time for needle-in and needle-out tasks.

“Appendix B” shows the minimum staffing requirements for each key stage for all three donor centres. Each graph depicts the average number of staff (rounded up), average staff utilisation, and the standard deviations (error bars) in each slot. The staff utilisation level is constantly high except at the start and end of the day. The high utilisation level suggests that the minimum staffing requirements found by the SBSA algorithm have provided near-optimal solutions for all three donor centres. Overall, the Townsville donor centre needs 85.5 total working hours, while the Gosford and Warrnambool donor centres need 75.5 and 57 total working hours, respectively, to serve all donors with an average waiting time of fewer than 6 min during a typical day.

7 Conclusions and future research

To summarise, we started this study by understanding Australian Red Cross Lifeblood donor centres’ donor arrival process, blood donation process, available resources, and sources of uncertainty in the donation process. Then, we discussed the importance of voluntary donors and the effect of donor satisfaction on the donor return rate. The key performance indicator, Time-to-Couch, is directly affected by long waiting times and, consequently, influences donor satisfaction. The demand for staff is inconsistent throughout the day, and fulfilling that demand ensures a smooth donor flow. Therefore, we introduced a simulation-based optimisation approach for determining the minimum staffing requirement in each key stage for a typical day.

First, we fitted probability distributions to data on the durations of registration, assessment, and questionnaire completion, as well as the collection of each blood product at all four donor centres. Second, a discrete event simulation model was developed to represent the entire functioning of a donor centre during a typical day, and it was used to obtain the average waiting times of each donor centre by changing the parameter values. Third, we modified a simulation-based simulated annealing algorithm for the Melbourne CBD donor centre together with a neighbour solution generation algorithm to determine the minimum staffing

requirement throughout the day. Next, we further investigated our method using different benchmarks from the Melbourne CBD donor centre. Last, we tested the applicability of our method by applying it to the Gosford, Townsville, and Warrnambool donor centres.

With regard to our contribution to the literature, no study exists that considers a complicated arrival process that is affected by appointment-based arrivals, earliness, tardiness, no-shows, and random arrivals. A key aspect of our method is that it avoids using an analytical model of the system, which may not only be too complicated but also analytically challenging. Therefore, we were able to develop a comprehensive simulation model to represent the entire donation process of a blood donor centre by considering all the sources of uncertainty. There are no studies that we could find that used a simulation-based simulated annealing algorithm for determining the minimum staffing requirement. To improve the efficiency of the simulated annealing algorithm, we introduced a novel method called the NS algorithm that seeks neighbouring staff configurations based on staff occupancy levels. Our method for detecting bottleneck/under-utilised staff slots using the *average full occupancy* levels plays a key role in this study. Therefore, our study filled many gaps in the literature and contributed novel approaches.

The existing data on registration and assessment durations includes some waiting time that is impossible to eliminate. Since the objective function of the simulated annealing algorithm does not include any service durations but only the average waiting time and total number of staff, we do not expect a significant deviation in the minimum staffing requirement from the actual one. To determine the minimum staffing requirement, we experimented with the most suitable parameters in the simulated annealing algorithm. However, in the process of calculating the average full occupancy of a given half-hour, we imposed an 80% threshold to indicate whether a half-hour is fully occupied or not on a given day, as per Equation (6). This occupancy threshold must be a parameter that can be adjusted appropriately, but we did not explore it due to a high computational expense in the simulation model. However, we believe 80% is a reasonable value according to the minimum staffing requirements determined in Sect. 5.

Any service centre can apply our method of calculating average full occupancy levels and utilising it to detect slots with a bottleneck or under-utilised staff. Since our simulation-based simulated annealing algorithm is computationally expensive, it is better to determine a more efficient way of generating the minimum staffing requirements. One possibility is applying a machine learning model such as neural networks (De Angelis et al., 2003).

Our model is very close to the real donor centre environment. However, we did not consider non-collection activities that staff perform, such as data entry and preparing machines. We did not consider donor adverse events, and donor deferrals in our simulation model due to their small probabilities. Recalling Sect. 1, these factors can be the main reasons for having skewed distributions for activities such as registration, assessment, and collection. Hence, it is not mandatory to consider them separately.

Most importantly, our next step is to determine the optimal shift schedule that satisfies the minimum staffing requirements determined by this study for each donor centre. For that, an integer programming problem can be developed that fulfills all Australian labour standards, which are slightly different from one state to another. This optimal shift schedule will maintain staff satisfaction by allocating rest and meal breaks to shifts while minimising staff non-utilised time.

The numerical results based on the four Australian Red Cross Life donor centres demonstrate that the proposed SBSA algorithm can be adapted to any donor centre to determine the minimum staffing requirement. Since these staffing demands ensure the donor waiting time target is met for each donor centre, they have the potential to improve donor satisfaction

and streamline the donor flow. The ability of this method to distribute the necessary number of employees at each stage will reduce work-related stress for employees and increase staff satisfaction. Furthermore, since we have minimised the total number of working hours, there is a potential to reduce costs. This study has been able to make a good contribution to filling gaps in the literature on the optimisation of time-dependent queueing systems. Finally, based on the overall outcomes, we can conclude that this method has the capacity to improve healthcare management in Australia because it offers a practical solution to efficiently reduce waiting times for donors and ensure a sustainable blood supply.

Acknowledgements We would like to appreciate the support and data received from Australian Red Cross Lifeblood. Australian governments fund Australian Red Cross Lifeblood to provide blood, blood products and services to the Australian community. Furthermore, this research project is financially supported by the University of Melbourne Research Scholarship. Furthermore, we would like to acknowledge the support of the Australian Research Council Industrial Transformation Training Centre in Optimisation Technologies, Integrated Methodologies, and Applications (OPTIMA) and the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Fitting probability distributions

Given the number of phases, we used the *mapfit* R package to fit Coxian phase-type distributions to the data. It returns information about the fitted distribution, including the number of EM algorithm iterations, the convergence status of the EM algorithm, the AIC, and the log-likelihood value. The AIC value decreases as the number of phases increases, and it begins to rise after a certain number of phases, after which the remaining phases are no longer tested. We chose the distribution that has the minimum AIC value as the best-fitted Coxian distribution. We followed the same process to select the number of components when fitting mixture distributions using *mixR* R package.

Kolmogorov-Smirnov (KS) test is a non-parametric test that is used to determine if a sample comes from a population with a particular distribution. In this study, the false-positive rate of the KS test is set at 0.01, and if the KS test's p -value is higher than that, we do not have sufficient data to reject the null hypothesis (the data follow a specified distribution).

A.1: Registration duration

See Table 23.

Table 23 The AIC, log-likelihood and the KS-test's p -values of the Coxian distributions fitted to registration duration data in each donor centre where the selected models are highlighted

# Phases	# Iterations	Log-likelihood	AIC	KS test p -value
<i>Melbourne CBD donor centre</i>				
35	427	-322,726.1	645,590.1	0.1076
36	430	-322,725.6	645,593.3	0.1070
37	650	-322,660.4	645,466.9	0.1300
38	639	-322,659.2	645,468.4	0.1305
<i>Gosford donor centre</i>				
7	966	-53,291.1	106,608.1	0.4079
8	1687	-53,287.1	106,604.3	0.7917
9	1484	-53,284.6	106,603.2	0.9140
10	270	-53,283.9	106,605.8	0.9140
<i>Townsville donor centre</i>				
6	1047	-18,589.9	37,201.9	0.0073
7	872	-18,567.4	37,160.8	0.2566
8	1233	-18,560.9	37,151.9	0.7669
9	1303	-18,560.4	37,154.8	0.7695
<i>Warrnambool donor centre</i>				
5	715	-15,354.7	30,727.4	0.0634
6	1867	-15,325.2	30,672.4	0.1128
7	3242	-15,320.9	30,667.9	0.1238
8	2021	-15,321.1	30,672.2	0.1246

A.2: Questionnaire completion duration

As shown in Tables 24 and 25 gamma, log-normal, Weibull, and Coxian phase-type distributions were unable to satisfy the KS test for the questionnaire completion duration data of the Melbourne CBD donor centre. Therefore, we selected the 98-phase Coxian distribution that has the smallest AIC value among all the distributions. Figure 14 shows a histogram of the empirical data on questionnaire completion duration at the Melbourne CBD donor centre overlaid with the 98-phase Coxian distribution. It indicates that the fit is poor. According to Okamura (2022), over-fitting can occur when *mapfit* package is used to estimate phase-type distribution parameters with high orders. The KS test might not be satisfied even though this 98-phase Coxian distribution was converged. The questionnaire completion duration data of the other three donor centres were well-fitted by Coxian phase-type distributions.

Table 24 Information about gamma, log-normal, and Weibull distributions fitted to questionnaire filling duration data in Melbourne CBD donor centre

Distribution	Parameter MLEs	AIC	BIC	KS test p -value
Gamma	$\hat{\alpha} = 1.1719$, $\hat{\beta} = 1.3485$	148,932.8	148,951.6	<0.0001
Log-normal	$\hat{\mu} = 1.7892$, $\hat{\sigma} = 0.6360$	675,826	675,845.4	<0.0001
Weibull	$\hat{\lambda} = 0.8804$, $\hat{k} = 1.0282$	150123.8	150142.6	<0.0001

Table 25 The AIC, log-likelihood and the KS-test's p -values of the Coxian distributions fitted to the questionnaire completion duration data in each donor centre where the selected models are highlighted

# Phases	# Iterations	Log-likelihood	AIC	KS test p -value
<i>Melbourne CBD donor centre</i>				
96	9221	-62,435.7	125,253.5	<0.0001
97	11,618	-62,431.9	125,249.9	<0.0001
98	10,753	-62,410.8	125,211.6	<0.0001
99	10,010	-62,410.2	125,214.4	<0.0001
<i>Gosford donor centre</i>				
26	8251	-22,895.1	45,892.2	0.0661
27	8332	-22,893.2	45,892.4	0.0854
28	11,130	-22,890.0	45,890.1	0.0942
29	8348	-22,889.1	45,892.2	0.1003
<i>Townsville donor centre</i>				
17	3290	-5412.4	10,890.9	0.6812
18	2993	-5411.1	10,892.2	0.6988
19	2969	-5404.3	10,882.6	0.7653
20	2804	-5402.9	10,883.8	0.7674
<i>Warrnambool donor centre</i>				
13	5520	-869.5	1788.9	0.1336
14	6048	-867.0	1788.1	0.1338
15	6502	-864.2	1786.2	0.1343
16	6629	-864.0	1790.0	0.1346

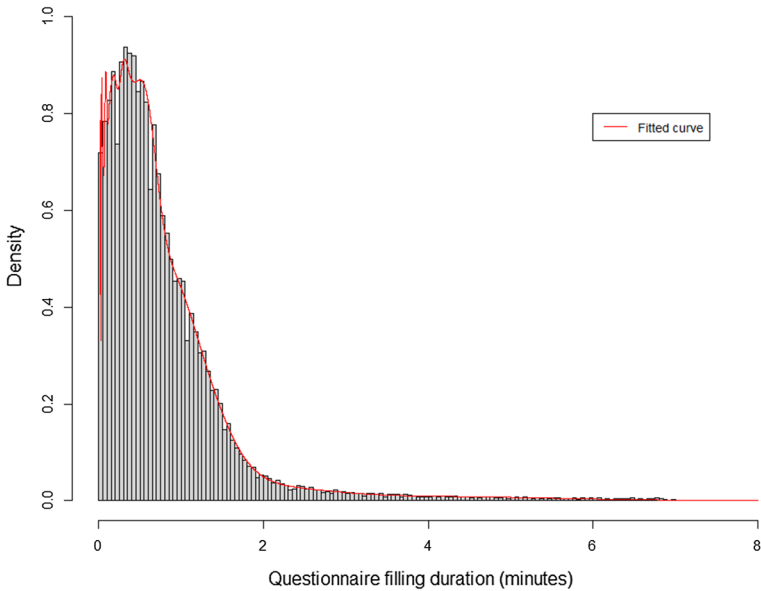


Fig. 14 Histogram and fitted 98-phase Coxian distribution for the data on questionnaire completion duration of the Melbourne CBD donor centre

A.3: Assessment duration

See Table 26.

Table 26 The AIC, log-likelihood and the KS-test's p -values of the Coxian distributions fitted to the assessment duration data in each donor centre where the selected models are highlighted

# Phases	# Iterations	Log-likelihood	AIC	KS test p -value
<i>Melbourne CBD donor centre</i>				
41	1711	-415,618.8	831,399.6	0.8772
42	984	-415,559.1	831,284.4	0.9867
43	952	-415,557.1	831,284.3	0.9875
44	918	-415,555.3	831,284.5	0.9875
<i>Gosford donor centre</i>				
16	1330	-59,956.2	119,974.4	0.6349
17	3249	-59,940.0	119,946.0	0.9955
18	2548	-59,939.6	119,949.3	0.9977
19	1433	-59,938.8	119,951.6	0.9985
<i>Townsville donor centre</i>				
9	4642	-23,934.5	47,903.1	0.4227
10	1427	-23,931.0	47,900.1	0.6254
11	647	-23,928.3	47,898.7	0.9908
12	1274	-23,928.3	47,902.7	0.9980
<i>Warrnambool donor centre</i>				
14	1840	-30,815.1	61,684.3	0.1235
15	3566	-30,804.8	61,667.6	0.2364
16	2209	-30,801.6	61,665.2	0.3853
17	1880	-30,800.6	61,667.1	0.4303

A.4: Whole blood collection duration

See Table 27.

Table 27 The number of components, AIC, BIC, log-likelihood and the KS-test's p -values of the mixture distributions fitted to the whole blood collection duration data in each donor centre where the selected models are highlighted

Distribution	# Components	Log-likelihood	AIC	BIC	p -value
<i>Melbourne CBD donor centre</i>					
Gamma mixture	4	-41,988.4	83,998.8	840,85.7	<0.001
Gamma mixture	5	-41,969.7	83,967.6	84,078.2	0.0879
Gamma mixture	6	-41,970.1	83,974.2	84,108.6	0.0879
Normal mixture	5	-41,981.5	83,990.9	84,101.6	0.06379
Weibull mixture	7	-41,476.0	82,992.0	83,150.0	0.0014
<i>Gosford donor centre</i>					
Gamma mixture	3	-17,904.7	35,825.5	35,882.7	< 0.001
Gamma mixture	4	-17,893.7	35,809.4	35,888.1	0.0750
Gamma mixture	5	-17,894.0	35,816.0	35,916.2	0.0321
Normal mixture	3	-17,916.2	35,848.6	35,905.8	0.1379
Weibull mixture	7	-17,513.2	35,066.5	35,209.7	0.0456
<i>Townsville donor centre</i>					
Gamma mixture	3	-783.1	1582.9	1620.6	0.0029
Gamma mixture	4	-772.8	1567.6	1610.9	0.1379
Gamma mixture	5	-768.4	1569.9	1626.3	0.1620
Normal mixture	4	-772.5	1567.1	1610.4	0.04556
Weibull mixture	3	-663.4	1342.9	1374.4	<0.0001
<i>Warrnambool donor centre</i>					
Gamma mixture	6	-12,419.4	24,872.7	24,987.2	<0.001
Gamma mixture	7	-12,411.8	24,863.7	24,998.3	0.075
Gamma mixture	8	-12,410.1	24,866.2	25,021.1	0.0638
Normal mixture	7	-10,486.9	21,013.8	21,148.5	0.0124
Weibull mixture	9	-10,978.0	22,008.1	22,183.2	<0.0001

A.5: Plasma collection duration

See Table 28.

Table 28 The number of components, AIC, BIC, log-likelihood and the KS-test's p -values of the mixture distributions fitted to the plasma collection duration data in each donor centre where the selected models are highlighted

Distribution	# Components	Log-likelihood	AIC	BIC	p -value
<i>Melbourne CBD donor centre</i>					
Normal mixture	5	-72,678.3	145,384.1	145,557.6	0.0010
Normal mixture	6	-72,635.1	145,304.2	145,438.6	0.1191
Normal mixture	7	-72,737.9	145,515.7	145,673.8	0.1379
Gamma mixture	13	-72,677.2	145,430.5	145,730.8	0.7823
Weibull mixture	11	-72,609.5	145,283.0	145,535.9	0.238
<i>Gosford donor centre</i>					
Normal mixture	6	-30,841.1	61,716.2	61,836.9	0.0268
Normal mixture	7	-30,826.3	61,692.6	61,834.5	0.0750
Normal mixture	8	-30,825.5	61,697.0	61,860.3	<0.001
Gamma mixture	8	-30,920.9	61,887.9	62,051.2	0.305
Weibull mixture	9	-30,855.5	61,763.8	61,947.7	0.0750
<i>Townsville donor centre</i>					
Normal mixture	4	-36,848.2	73,692.1	73,800.2	<0.001
Normal mixture	5	-36,823.5	73,675.2	73,776.3	0.3431
Normal mixture	6	-36,811.1	73,740.5	73,791.5	0.4002
Gamma mixture	5	-36,875.4	73,778.9	73,879.9	0.305
Weibull mixture	4	-36,862.3	73,746.6	73,826.1	0.1191
<i>Warrnambool donor centre</i>					
Normal mixture	2	-22,124.7	44,024.8	44,071.1	<0.001
Normal mixture	3	-21,995.0	44,006.1	44,059.7	0.2700
Normal mixture	2	-21,900.3	44,030.4	44,099.6	0.3017
Gamma mixture	3	-22,006.2	44,028.5	44,082.1	0.7319
Weibull mixture	10	-21,910.0	43,878.1	44,072.6	0.4282

A.6: Single-dose platelets collection duration

See Table 29.

Table 29 The number of components, AIC, BIC, log-likelihood and the KS-test's p -values of the mixture distributions fitted to single-dose platelets collection duration data in Melbourne CBD donor centres where the selected models are highlighted

Distribution	# Components	Log-likelihood	AIC	BIC	p -value
Gamma mixture	4	-3044.8	6111.6	6163.4	<0.001
Gamma mixture	5	-3039.5	6107.0	6172.9	0.4282
Gamma mixture	6	-3032.0	6098.0	6178.1	0.5089
Normal mixture	5	-3041.2	6110.6	6176.5	0.1191
Weibull mixture	13	-3003.4	6082.9	6261.8	0.1025

A.7: Double-dose platelets collection duration

See Table 30.

Table 30 The number of components, AIC, BIC, log-likelihood and the KS-test’s *p*-values of the mixture distributions fitted to double-dose platelets collection duration data in the Melbourne CBD donor centres where the selected models are highlighted

Distribution	# Components	Log-likelihood	AIC	BIC	<i>p</i> -value
Normal mixture	4	−8600.8	17,223.7	17,285.5	0.1826
Normal mixture	5	−8578.7	17,185.3	17,264.1	0.238
Normal mixture	6	−8590.1	17,214.2	17,309.8	0.3063
Gamma mixture	4	−8622.1	17,266.2	17,328.1	0.3431
Weibull mixture	8	−8590.0	17,226.1	17,355.3	0.3842

Appendix B: Minimum staffing requirements

B.1: Gosford donor centre

See Figs. 15, 16 and 17.

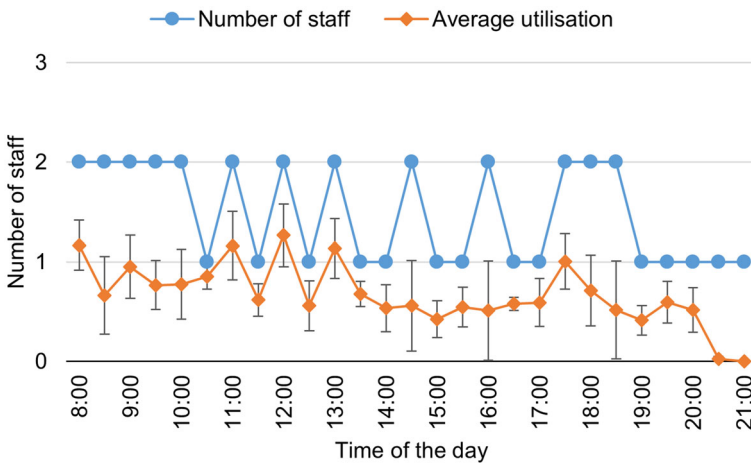


Fig. 15 Minimum staffing requirement with their utilisation over the day—registration stage

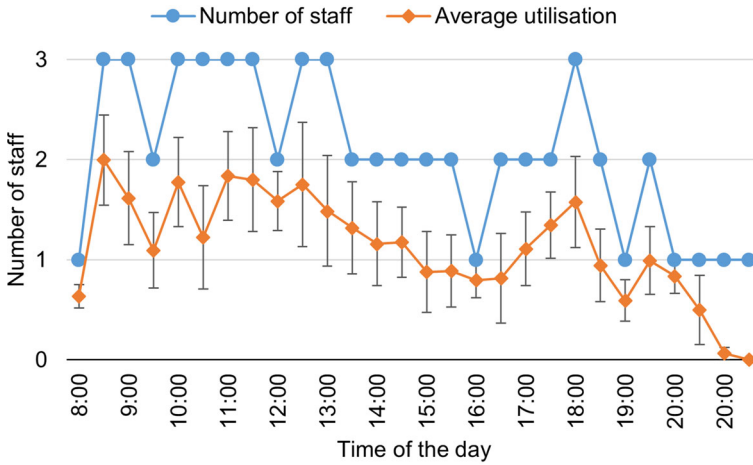


Fig. 16 Minimum staffing requirement with their utilisation over the day—assessment stage

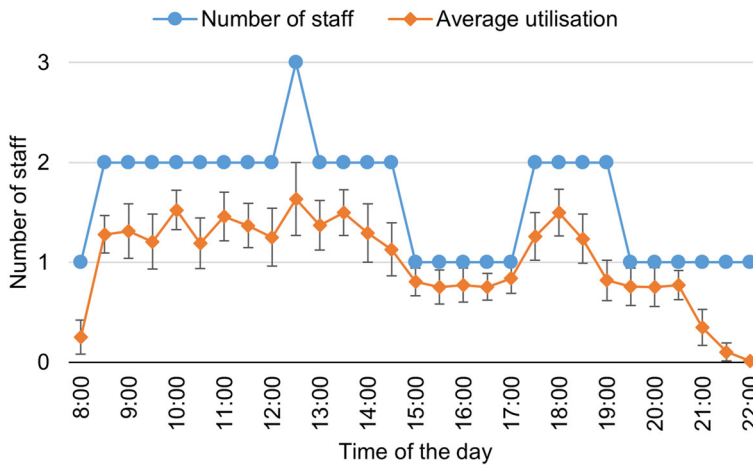


Fig. 17 Minimum staffing requirement with their utilisation over the day—collection stage

B.2: Townsville donor centre

See Figs. 18, 19 and 20.

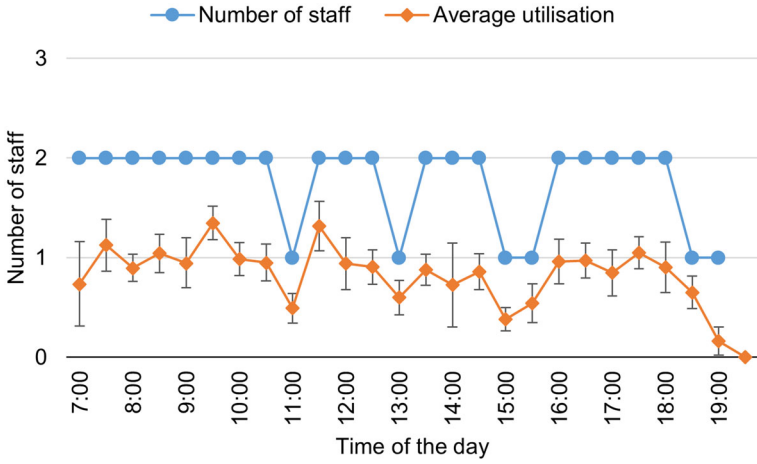


Fig. 18 Minimum staffing requirement with their utilisation over the day—registration stage

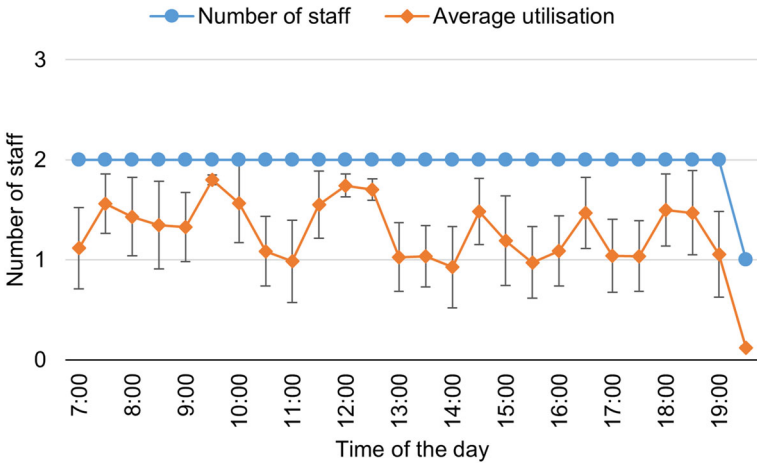


Fig. 19 Minimum staffing requirement with their utilisation over the day—assessment stage

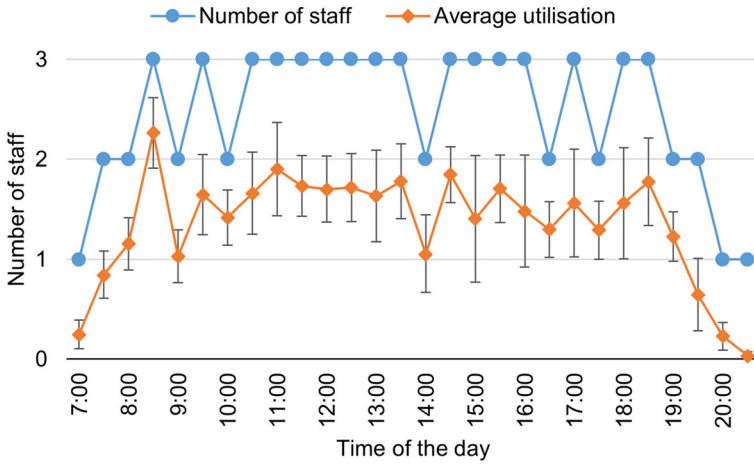


Fig. 20 Minimum staffing requirement with their utilisation over the day—collection stage

Warrnambool donor centre

See Figs. 21, 22 and 23.

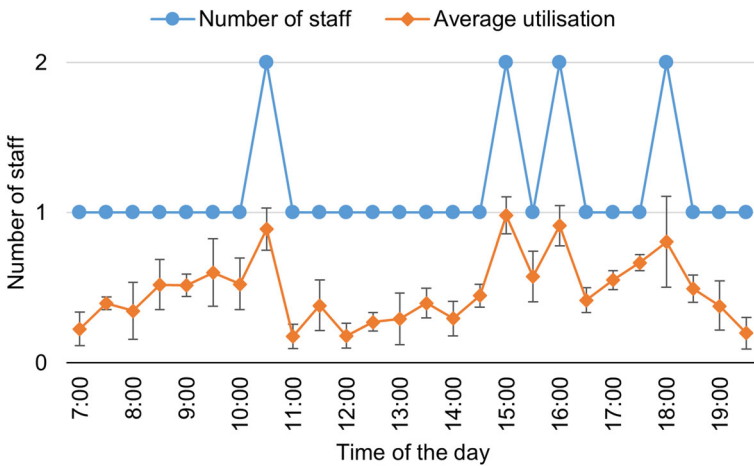


Fig. 21 Minimum staffing requirement with their utilisation over the day—registration stage

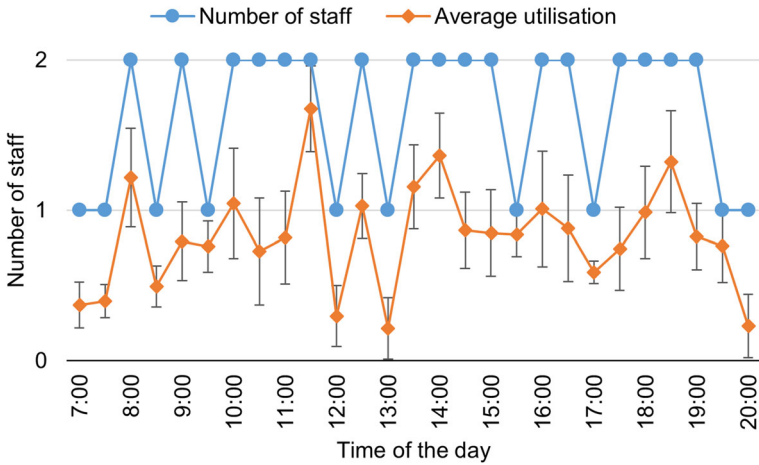


Fig. 22 Minimum staffing requirement with their utilisation over the day—assessment stage

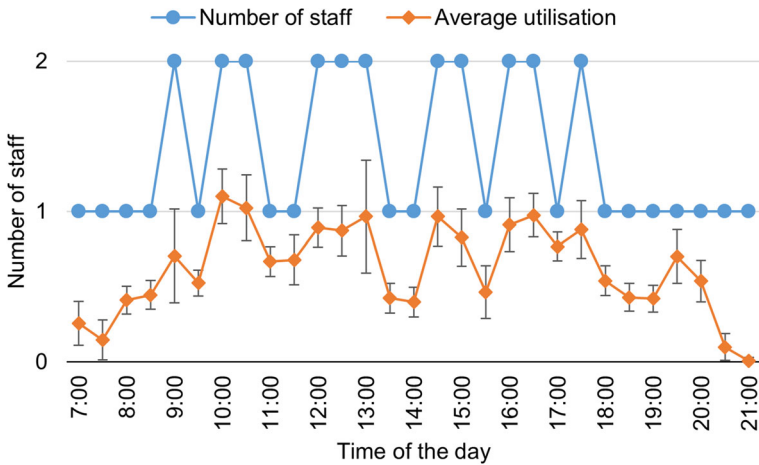


Fig. 23 Minimum staffing requirement with their utilisation over the day—collection stage

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the 2nd international symposium on information theory* (pp. 267–281). Akademiai Kiado.
- Ahmed, M. A., & Alkhamis, T. M. (2002). Simulation-based optimisation using simulated annealing with ranking and selection. *Computers and Operations Research*, 29(4), 387–402.
- Ahmed, M. A. (2007). A modification of the simulated annealing algorithm for discrete stochastic optimisation. *Engineering Optimisation*, 39(6), 701–714. <https://doi.org/10.1080/03052150701280533>
- Ahmed, M. A., & Alkhamis, T. M. (2009). Simulation optimisation for an emergency department health care unit in Kuwait. *European Journal of Operational Research*, 198, 936–942. <https://doi.org/10.1016/j.ejor.2008.10.025>
- Alfonso, E., Xie, X., & Augusto, V. (2015). A simulation-optimisation approach for capacity planning and appointment scheduling of blood donors based on mathematical programming representation of event dynamics. In *2015 IEEE international conference on automation science and engineering (CASE)* (pp. 728–733). <https://ieeexplore.ieee.org/document/7294167>.

- Amaran, S., Sahinidis, N. V., Sharda, B., & Bury, S. J. (2016). Simulation optimisation: A review of algorithms and applications. *Annals of Operations Research*, 240(1), 351–380. <https://doi.org/10.1007/s10479-015-2019-x>
- ARCL: Australian Red Cross Lifeblood. Retrieved January 9, 2022, from <https://www.lifeblood.com.au/about/who-we-are>
- Asmussen, S., Nerman, O., & Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 419–441. https://www.researchgate.net/publication/238974468_Fitting_Phase-Type_Distributions_via_the_EM_Algorithm
- Baker, K. R., Burns, R. N., & Carter, M. (1979). Staff scheduling with day-off and workstretch constraints. *AIIE Transactions*, 11, 286–292. <https://doi.org/10.1080/05695557908974473>
- Bayram, H., & Sahin, R. (2013). A new simulated annealing approach for traveling salesman problem. *Mathematical and Computational Applications*, 18(3), 313–322. <https://doi.org/10.3390/mca18030313>
- Blake, J. T., & Shimpla, S. (2014). Determining staffing requirements for blood donor clinics: The Canadian blood services experience. *Transfusion*, 54, 814–820. <https://doi.org/10.1111/trf.12353>
- Bulgak, A. A., & Sanders, J. L. (1988). Integrating a modified simulated annealing algorithm with the simulation of a manufacturing system to optimise buffer sizes in automatic assembly systems. In *1988 winter simulation conference proceedings* (pp. 684–690). <https://ieeexplore.ieee.org/document/716241>
- Connor, A. M., & Shea, K. (2000). A comparison of semi-deterministic and stochastic search techniques. In I. C. Parmee (Ed.), *Evolutionary Design and Manufacture*. Springer. https://doi.org/10.1007/978-1-4471-0519-0_23
- Cox, D. R. (1955). A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(2), 313–319. <https://doi.org/10.1017/S0305004100030231>
- De Angelis, V., Felici, G., & Impelluso, P. (2003). Integrating simulation and optimisation in health care centre management. *European Journal of Operational Research*, 150(1), 101–114. [https://doi.org/10.1016/S0377-2217\(02\)00791-9](https://doi.org/10.1016/S0377-2217(02)00791-9)
- Defraeye, M., & Van Nieuwenhuysse, I. (2016). Staffing and scheduling under nonstationary demand for service: A literature review. *Omega*, 58, 4–25. <https://doi.org/10.1016/j.omega.2015.04.002>
- Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4), 1–34.
- Devine, D., Goldman, M., Engelfriet, C. P., Reesink, H. W., Hetherington, C., Hall, S., Steed, A., Harding, S., Westman, P., Gogarty, G., Katz, L. M., & Bryant, M. (2007). Donor recruitment research. *Vox Sanguinis*, 93, 250–259. <https://doi.org/10.1111/j.1423-0410.2007.00962.x>
- Donnelly, C., McPetridge, L. M., Marshall, A. H., & Mitchell, H. J. (2018). A two-stage approach to the joint analysis of longitudinal and survival data utilising the Coxian phase-type distribution. *Statistical Methods in Medical Research*, 27(12), 3577–3594. <https://doi.org/10.1177/0962280217706727>
- El-Rifai, O., Garaix, T., Augusto, V., & Xie, X. (2015). A stochastic optimisation model for shift scheduling in emergency departments. *Healthcare Management Science*, 18(3), 289–302. <https://doi.org/10.1007/s10729-014-9300-4>
- Ernst, A. T., Jiang, H., Krishnamoorthy, M., & Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research*, 153(1), 3–27. [https://doi.org/10.1016/S0377-2217\(03\)00095-X](https://doi.org/10.1016/S0377-2217(03)00095-X)
- Fackrell, M. (2009). Modelling healthcare systems with phase-type distributions. *Health Care Management Science*, 12(11), 11–26. <https://doi.org/10.1007/s10729-008-9070-y>
- Faddy, M., Graves, N., & Pettitt, A. (2009). Modeling length of stay in hospital and other right skewed data: Comparison of phase-type, gamma and log-normal distributions. *Value in Health: Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 12(2), 309–314. <https://doi.org/10.1111/j.1524-4733.2008.00421.x>
- Frank, J. M., Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253), 68–78. <https://doi.org/10.1080/01621459.1951.10500769>
- Gardiner, J., Luo, Z., Tang, X., Tang, X., & Ramamoorthi, R. V. (2014). Fitting heavy-tailed distributions to health care data by parametric and bayesian methods. *Journal of Statistical Theory and Practice*, 8(4), 619–652. <https://doi.org/10.1080/15598608.2013.824823>
- Gunawan, A., & Lau, H. C. (2013). Master physician scheduling problem. *The Journal of the Operational Research Society*, 64(3), 410–425. <https://doi.org/10.1057/jors.2012.48>
- Henderson, D., Jacobson, S., & Johnson, A. (2003). The theory and practice of simulated annealing. *Handbook of Metaheuristics*. https://doi.org/10.1007/0-306-48056-5_10
- Jog, P., Suh, J. Y., & Van Gucht, D. (1991). Parallel genetic algorithms applied to the traveling salesman problem. *SIAM Journal on Optimization*, 1(4), 515–529. <https://doi.org/10.1137/0801031>

- Jorge, H., & John, M. (1992). Simulation optimisation using simulated annealing. *Computers and Industrial Engineering*, 22(4), 387–395. [https://doi.org/10.1016/0360-8352\(92\)90014-B](https://doi.org/10.1016/0360-8352(92)90014-B)
- Kamboj, M. S., & Sengupta, J. (2009). Comparative analysis of simulated annealing and tabu search channel allocation algorithms. *International Journal of Computer Theory and Engineering*, 588–591.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimisation by simulated annealing. *Science*, 220(4598), 671–680. <https://doi.org/10.1126/science.220.4598.671>
- Kriegeskorte, N. (2015). Crossvalidation. *Brain Mapping*. <https://doi.org/10.1016/B978-0-12-397025-1.00344-4>
- Kumar, A., & Anjomshoa, H. (2019). A two-stage model to predict surgical patients' lengths of stay from an electronic patient database. *IEEE Journal of Biomedical and Health Informatics*, 23(2), 848–856. <https://doi.org/10.1109/JBHI.2018.2819646>
- Kumar, A. (2019). *Intelligent Management of Elective Surgery Patient Flow*. PhD thesis, The University of Melbourne. <https://doi.org/10.13140/RG.2.2.10142.28486>.
- Lafayette, L., Sauter, G., Vu, L., & Meade, B. (2016). Spartan performance and flexibility: An HPC-Cloud Chimera. *OpenStack Summit, Barcelona*. <https://doi.org/10.4225/49/58ead90dceaa>
- Latouche, G., & Ramaswami, V. (1993). A logarithmic reduction algorithm for quasi-birth-death processes. *Journal of Applied Probability*, 30(3), 650–674. <https://doi.org/10.2307/3214773>
- Lewis, P. A. W., & Shedler, G. S. (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3), 403–413. <https://doi.org/10.1002/nav.3800260304>
- Lovelace-Tozer, M. S. (2018). Analysing and modelling donor flow in an Australian red cross blood service donation centre. Master's thesis, The University of Melbourne.
- Marazzi, A., Paccaud, F., Ruffieux, C., & Beguin, C. (1998). Fitting the distributions of length of stay by parametric models. *Medical Care*, 36(6), 915–927.
- Masser, B. M., Davison, T. E., & Chapman, C. M. (2016). How can we encourage our voluntary non-remunerated donors to donate more frequently? *VOXS*, 12, 112–118. <https://doi.org/10.1111/voxs.12312>
- Matsuo, H., Suh, C. J., & Sullivan, R. S. (1989). A controlled search simulated annealing method for the single machine weighted tardiness problem. *Annals of Operations Research*, 21(1), 85–108. <https://doi.org/10.1007/BF02022094>
- McClellan, S., & Millard, P. (1993). Patterns of length of stay after admission in geriatric medicine: An event history approach. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 42(3), 263–274. <https://doi.org/10.2307/2348804>
- Mihaylova, B., Briggs, A., O'Hagan, A., & Thompson, S. G. (2011). Review of statistical methods for analysing healthcare resources and costs. *Health Economics*, 20(8), 897–916. <https://doi.org/10.1002/hec.1653>
- Nasr, W. W., & Taaffe, M. R. (2013). Fitting the Ph/Mt/s/c time-dependent departure process for use in tandem queueing networks. *INFORMS Journal on Computing*, 25(4), 758–773. <https://doi.org/10.1287/ijoc.1120.0538>
- Neuts, M. F. (1975). Computational uses of the method of phases in the theory of queues. *Computers & Mathematics with Applications*, 1(2), 151–166. [https://doi.org/10.1016/0898-1221\(75\)90015-2](https://doi.org/10.1016/0898-1221(75)90015-2)
- Okamura, H. (2022, May 31). *mapfit*. GitHub. Retrieved August 8, 2022, from <https://github.com/okamumu/mapfit>
- Okamura, H., & Dohi, T. (2016). Fitting phase-type distributions and Markovian arrival processes: Algorithms and tools. In: L. Fiondella, & A. Puliafito (Eds.), *Principles of performance and reliability modelling and evaluation* (pp. 49–75). Springer Series in Reliability Engineering. Springer, Cham. https://doi.org/10.1007/978-3-319-30599-8_3
- Pflug, G. (2009). Discrete stochastic optimisation. *Encyclopedia of Optimisation*. https://doi.org/10.1007/978-0-387-74759-0_132
- Rodriguez, C., Garaix, T., Xie, X., & Augusto, V. (2015). Staff dimensioning in homecare services with uncertain demands. *International Journal of Production Research*, 53(24), 7396–7410. <https://doi.org/10.1080/00207543.2015.1081427>
- Rohleder, T. R., Lewkonja, P., Bischak, D. P., Duffy, P., & Hendijani, R. (2011). Using simulation modelling to improve patient flow at an outpatient orthopedic clinic. *Health Care Management Science*, 14(2), 135–145. <https://doi.org/10.1007/s10729-010-9145-4>
- Ross, S. (2013). *Simulation: Elements of Probability*. (5th Ed., pp.5–38) Academic Press. <https://doi.org/10.1016/B978-0-12-415825-2.00002-4>
- Shi-hua, Z., Sannuo, L., Ze-jun, Z., & Yiwen, Z. (2016). List-based simulated annealing algorithm for traveling salesman problem. *Computational Intelligence and Neuroscience*. <https://doi.org/10.1155/2016/1712630>
- Siegel, A.F. & Wagner, M.R. (2022). Chapter 3 - histograms: Looking at the distribution of data. In Andrew F. Siegel & Michael R. Wagner (Eds.), *Practical Business Statistics* (8th ed., pp. 45–74). Academic Press. <https://www.sciencedirect.com/science/article/pii/B9780128200254000038>.

- Smith, J. S., & Nelson, B. L. (2015). Estimating and interpreting the waiting time for customers arriving to a non-stationary queueing system. *Winter Simulation Conference (WSC), 2015*, 2610–2621. <https://doi.org/10.1109/WSC.2015.7408369>
- Suman, B., & Kumar, P. (2006). A survey of simulated annealing as a tool for single and multiobjective optimisation. *The Journal of the Operational Research Society*, 57(10), 1143–1160.
- Van Brummelen, S. P. J., Van Dijk, N. M., Van den Hurk, K., & De Kort, W. L. (2018). Waiting time-based staff capacity and shift planning at blood collection sites. *Health Systems*, 7(2), 89–99. <https://doi.org/10.1057/s41306-017-0032-9>
- Wang, J., Quan, S., Li, J., & Hollis, A. M. (2012). Modelling and analysis of workflow and staffing level in a computed tomography division of University of Wisconsin medical foundation. *Health Care Management Science*, 15(2), 108–120. <https://doi.org/10.1007/s10729-011-9188-1>
- Weglarczyk, S. (2018). Kernel density estimation and its application. *ITM Web of Conferences*, 23, EDP Sciences. <https://doi.org/10.1051/itmconf/20182300037>
- Williford, E., Haley, V., McNutt, L. A., & Lazariu, V. (2020). Dealing with highly skewed hospital length of stay distributions: The use of gamma mixture models to study delivery hospitalisations. *PLOS one*. <https://doi.org/10.1371/journal.pone.0231825>
- Yu, Y. (2022). mixR: An R package for finite mixture modeling for both raw and binned data. *Journal of Open Source Software*. <https://doi.org/10.21105/joss.04031>
- Zhou, A. H., Zhu, L. P., Hu, B., Deng, S., Song, Y., Qiu, H., & Pan, S. (2015). Traveling-salesman-problem algorithm based on simulated annealing and gene-expression programming. *Information*. <https://doi.org/10.3390/info10010007>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.