Demand forecasting for platelet usage: From univariate time series to multivariable models

Maryam Motamedi¹, Jessica Dawson¹, Na Li¹,²,³, Douglas G. Down¹*, Nancy M. Heddle³,⁴

¹ Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada, ² Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada, ³ Michael G. DeGroote Centre for Transfusion Research, Faculty of Health Sciences, Hamilton, Ontario, Canada, ⁴ Centre for Innovation, Canadian Blood Services, Ottawa, Ontario, Canada

* downd@mcmaster.ca

Abstract

Platelet products are both expensive and have very short shelf lives. As usage rates for platelets are highly variable, the effective management of platelet demand and supply is very important yet challenging. The primary goal of this paper is to present an efficient forecasting model for platelet demand at Canadian Blood Services (CBS). To accomplish this goal, five different demand forecasting methods, ARIMA (Auto Regressive Integrated Moving Average), Prophet, lasso regression (least absolute shrinkage and selection operator), random forest, and LSTM (Long Short-Term Memory) networks are utilized and evaluated via a rolling window method. We use a large clinical dataset for a centralized blood distribution centre for four hospitals in Hamilton, Ontario, spanning from 2010 to 2018 and consisting of daily platelet transfusions along with information such as the product specifications, the recipients’ characteristics, and the recipients’ laboratory test results. This study is the first to utilize different methods from statistical time series models to data-driven regression and machine learning techniques for platelet transfusion using clinical predictors and with different amounts of data. We find that the multivariable approaches have the highest accuracy in general, however, if sufficient data are available, a simpler time series approach appears to be sufficient. We also comment on the approach to choose predictors for the multivariable models.

1 Introduction

Platelet products are a vital component of patient treatment for bleeding problems, cancer, AIDS, hepatitis, kidney or liver diseases, traumatology and in surgeries such as cardiovascular surgery and organ transplants [1]. In addition, miscellaneous platelet usage and supply are associated with several factors such as patients with severe bleeding, trauma patients, aging population and emergence of a pandemic like COVID-19 [2]. The first two factors affect the uncertain demand pattern, while the latter two factors result in donor reduction. Platelet products have five to seven days shelf life before considering test and screening processes that
typically last two days [3], so the remaining shelf life of the platelets that arrive at the hospitals is typically between three to five days. The extremely short shelf life along with the highly variable daily platelet usage makes platelet demand and supply management a highly challenging task, invoking a robust blood product demand and supply system.

Canadian Blood Services (CBS) is the national blood supplier for Canadian patients. The current blood supply chain for CBS is an integrated network consisting of a regional CBS distribution centre and several hospitals, as illustrated in Fig 1. Hospitals request blood products from the regional blood centres for the next day, yet the regional blood centres are not aware of the actual demands as each hospital has its own blood bank. Furthermore, recipient demographics and hospital inventory management systems are not disclosed to CBS or the regional blood centres. Hospitals hold excess inventory to manage the highly variable platelet demand. However, holding surplus inventory makes platelet demand forecasting even more challenging for blood distribution centres. In particular, it results in wastage and does not allow the blood suppliers to recognize the real demand, which in turn yields an inefficient demand forecasting system. Accordingly, accurately forecasting the demand for blood products is a core requirement of a robust blood demand and supply management system.

This research is motivated by the platelet management problem confronted by CBS. Currently, there is a yearly wastage rate of about 9% for hospitals in Hamilton, Ontario (with an approximate cost of $400,000 per year) and about 15% for CBS with seasonal variation [4]. The current frequent same-day urgent orders, considered as shortages, are about 14% of the total orders in Hamilton, Ontario. Given the wastage rates and shortage rates, forecasting short-term demand for platelets is of particular value. In this research, we forecast platelet demand to overcome the mentioned challenges. The forecasting models used in this research can help both suppliers and consumers of platelets to make operational decisions, including inventory decisions, by providing information about future demand.

Fig 1. CBS blood supply chain with one regional blood centre and multiple hospitals.

https://doi.org/10.1371/journal.pone.0297391.g001
We study a large clinical database with 61377 platelet transfusions for 47496 patients in hospitals in Hamilton, Ontario from 2010 to 2018. We analyze the database to extract trends and patterns and find relations between the demand and clinical predictors. We find that there are three key issues that should be considered in the demand forecasting process: seasonality, the effect of clinical predictors on demand, and nonlinear dependencies among these clinical predictors. Consequently, we progressively build five demand forecasting models (of increasing complexity) that address these issues. The proposed methods are applied on the data to determine the influence of demand history as well as clinical predictors on demand forecasting. The first two models are univariate time series that only consider the demand history, while the remaining three methods are multivariable forecasting models that consider clinical predictors. Multivariable models are able to handle multiple parallel input sequences (time series). We explore the value gained from including a range of clinical predictors for platelet demand forecast models. More specifically, we consider clinical predictors consisting of laboratory test results, patient characteristics, and hospital census data as well as operational-related indicators, including the previous week’s platelet usage and the previous day’s received units, with the aim of accurate demand forecasting. In addition to the linear effects of the clinical predictors, we study the nonlinear effects of these clinical predictors in our choice of machine learning models. Moreover, we explore the effect of having different amounts of data on the accuracy of the forecasting methods. To the best of our knowledge, this study is the first that utilizes and evaluates different demand forecasting methodologies from univariate time series to multivariable models for platelet products and explores the effect of the amount of available data on these approaches.

The rest of this paper is organized as follows. In Section 2 we provide a literature review of demand forecasting methods for blood products, with a focus on platelets. Section 3 provides the data description and an overview of the five models used for forecasting platelet demand. The main results of the study are presented in Section 4. In Section 5, a comparison of the models is provided, and finally, in Section 6, concluding remarks are provided, including a discussion of ongoing work for this problem.

2 Literature review

2.1 Impact of blood demand forecasting on Blood Supply Chain inventory management policies

A Blood Supply Chain (BSC) consists of the collection, testing, production, and distribution of blood from donors to patients. These patients can be either routine (regular) patients who require blood products as a part of their treatment or emergency patients such as patients with the need of a blood product for surgeries or emergency trauma treatment. Therefore, there is a vital need to have a sufficient supply of blood products; otherwise, loss of lives may result. The structure of blood supply chains may differ from country to country, but generally, a blood supply chain consists of three main facilities (levels), including collection (donation) centres, blood centres, and hospitals. Moreover, there are various aspects to a blood supply chain such as designing the system, decision-making about the process such as collection, transportation and inventory management, and forecasting. One of the main aspects of a BSC is the ordering decisions and inventory management of blood products. Cohen and Pierskalla (1979) [5] is one of the earliest works on ordering policies for the red blood cell supply chain. They show that the optimal ordering policy is a base-stock value that depends on daily demand, the average transfusion to cross-match ratio, and the cross-match release period. Subsequently, an extensive body of research is dedicated to the inventory management of blood products [6–9].
A significant portion of existing studies focuses on formulating ordering decisions under the assumption of an independent and identically distributed sequence of demands. However, given the abundance of data available today, it is possible to construct demand forecasts that can be incorporated into the inventory system as additional demand information for determining ordering quantities. Many organizations perform demand forecasting as part of their decision-making processes, and incorporating these forecasts into the inventory model can be beneficial. Thus, when demand data is available, one can benefit from including additional information in the inventory management process. There is a recent stream of studies that incorporate additional demand information in the inventory management process. These studies can be categorized into two main groups: (i) prediction and optimization as a single step, (ii) predict then optimize.

In the first group, demand forecasts are included in the inventory optimization problem, rather than being a separate process. These models include additional demand information in the inventory model indirectly. In other words, there is no separate process for forecasting the demand, and demand is predicted inside the inventory model. Guan et al. (2017) [10] propose a convex optimization problem in which they forecast the platelet demand for several days into the future and build an optimal ordering policy based on the predicted demand, concentrating on minimizing the wastage while maintaining a minimum inventory level. Closely related to the work of Guan et al. (2017) [10], Abouee-Mehrizi et al. (2022) [11] consider a periodic review, perishable inventory control problem over a finite horizon, with zero lead-time and propose two models, a fixed age model and a robust model, for platelet inventory management. The objective is to determine daily ordering quantities while minimizing the wastage and shortage costs over this finite horizon. Demand is satisfied according to the Oldest-Unit-First-Out (OUFO) allocation policy, and unsatisfied demand is lost. The demand forecasts are included in the inventory optimization problem, rather than being a separate process. These models include additional demand information in the inventory model indirectly. In other words, there is no separate process for forecasting the demand, and demand is predicted inside the inventory model.

The second group that utilize data in the inventory model follow a two step process of first forecasting the demand and next using demand forecasts for optimizing inventory decisions. A classical example of this approach is presented in Elmachtoub and Grigas (2022) [12]. Li et al. 2021 [13] propose a data-driven multi-period inventory problem for RBC products that includes RBC demand predictions. They forecast the RBC demand and incorporate the forecasts in the inventory model. Since forecast errors exist, they introduce two extra decision variables, target inventory and reorder constraints, to control these errors in the ordering policies. Their model is a variation of the classical $(s, S)$ policy in which they define $S$ to compensate for demand overestimations, and define $s$ to compensate for demand underestimations. Both $s$ and $S$ are calculated based on the data and the predicted demand. Closely related to this stream, we consider a separate process for demand forecasting. We believe this allows suppliers to strategically make decisions in various parts of the supply chain, such as production planning and resource allocation.

2.2 Forecasting methods in Blood Supply Chain

There is a limited literature on platelet demand forecasting; most investigate univariate time series methods. In these studies, forecasts are based solely on previous demand values, without considering other features that may affect the demand. Critchfield et al. [14] develop models for forecasting platelet usage in a blood centre using several time series methods including Moving Average (MA), Winter’s method, and Exponential Smoothing (ES). Silva Filho et al.
develop a Box-Jenkins Seasonal Autoregressive Integrated Moving Average (BJ-SARIMA) model to forecast weekly demand for blood components, including platelets, in hospitals. They later extend this work in [16]. Kumari and Wijayanayake [17] propose a blood inventory management model for the daily supply of platelets focusing on reducing platelet shortages by applying three time series methods, MA, Weighted Moving Average (WMA) and ES. Volken et al. [18] use generalized additive regression and time-series models with ES to predict future whole blood donation, including platelets, and RBC transfusion trends. Fanooodi et al. [19] use artificial neural networks and ARIMA models to forecast platelet demand by considering daily demands for eight types of blood platelets. They consider different demand data lags, 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 15 days, 30 days, 90 days, 120 days, and 365 days, as the input data for the artificial neural networks.

On the other hand, many studies focus on univariate whole blood demand forecasting rather than a specific blood product, using time series or machine learning models. Frankfurter et al. [20] develop transfusion forecasting models using ES methods for a blood collection and distribution centre. Fortsch and Khapalova [21] apply various approaches to predict blood demand such as Naïve, MA, ES, and multiplicative Time Series Decomposition (TSD), amongst which a Box-Jenkins (ARMA) approach results in the highest prediction accuracy. Lestari et al. [22] apply four time series models, MA, WMA, ES and ES with trend, to forecast demand for blood components. Twumasi and Twumasi [23] apply K-Nearest Neighbour regression (KNN), Generalised Regression Neural Network (GRNN), Neural Network Auto-regressive (NNAR), Multi-Layer Perceptron (MLP), Extreme Learning Machine (ELM), and an LSTM network for forecasting and backcasting blood demand to predict future and lost past demand data respectively, by using a rolling-origin procedure.

Several recent studies include additional features other than demand history for demand forecasting. Drackley et al. [24] estimate long-term blood demand for Ontario, Canada based on previous transfusions’ age and sex-specific patterns. They forecast blood supply and demand for Ontario by considering demand and supply patterns, and demographic forecasts, with the assumption of fixed patterns and rates over time. Khalidi et al. [25] apply Artificial Neural Networks (ANNs) to forecast the monthly demand for three blood components, red blood cells (RBCs), platelets, and plasma for a case study in Morocco. Guan et al. [10] propose an optimization ordering strategy in which they forecast the platelet demand for several days into the future and build an optimal ordering policy based on the predicted demand by considering data features that affect the platelet demand. As mentioned in the Section 2.1, they integrate their demand model in the inventory management problem. Li et al. [13] develop a hybrid model consisting of seasonal and trend decomposition using Loess (STL) time series and eXtreme Gradient Boosting (XGBoost) for RBC demand forecasting.

2.3 Research gap and contributions

The blood demand forecasting literature is mainly focused on univariate demand forecasting. A few studies incorporate additional features beyond demand history, but they do not delve into the impact of these additional features on demand forecasts. In this study, we utilize multiple demand forecasting methods, including univariate analysis (time series methods) and multivariable analysis (regression and machine learning methods), and evaluate the performance of these models for platelet demand forecasting. In contrast to current works, multiple forecasting models are utilized to pursue the following goals: i) platelet demand forecasting for the benefit of both CBS and hospitals, ii) investigating the impact of clinical predictors on the
platelet demand, iii) exploring the effect of different amounts of data on the forecasting models; and iv) reducing the bullwhip effect, as a consequence of effective demand forecasting. The main contributions of this study are as follows:

1. We analyze the time series of platelet transfusion data by decomposing it into trend, seasonality and residuals, and detect meaningful patterns such as weekday/weekend and holiday effects that should be considered in any platelet demand predictor.

2. We utilize five different demand forecasting methods from univariate time series methods to multivariable methods including regression and machine learning. Since CBS has no access to recipient demographic data, our first method, ARIMA, used as the baseline for comparison, only considers demand history for forecasting, while the second model, Prophet, includes seasonalities, trend changes and holiday effects. We found that these models have issues with respect to accuracy, in particular when a limited amount of data are available, accordingly we apply a lasso regression method to include clinical predictors for demand forecasting. Finally, random forests and LSTM networks are used for demand forecasting to explore the nonlinear dependencies among the clinical predictors and the demand.

3. We utilize clinical predictors in the demand forecasting process, and select those that are most impactful by using lasso regression for structural variable selection and regularization. Results show that incorporating the clinical predictors in demand forecasting enhances forecasting accuracy.

4. We investigate the effect of different amounts of data on the forecasting accuracy and model performance and provide a holistic evaluation and comparison for different forecasting methods to evaluate the effectiveness of these models for different data types, providing suggestions on using these robust demand forecasting strategies in different circumstances. Results show that when having a limited amount of data (two years in our case), multivariable models outperform the univariate models, whereas having a large amount of data (eight years in our case) results in the ARIMA model performing nearly as well as the multivariable methods.

3 Methods

In this section, we present the general problem setting, a comprehensive data description and the methods used for data exploration. We also provide an overview of the five models used for forecasting platelet demand, an overview of the rolling window analysis used for retraining the models, and the error measures used for evaluation.

3.1 Problem setting

In this study, we consider a blood supply system consisting of one regional CBS distribution centre and four major hospitals operating in the city of Hamilton, Ontario. As a result of internal inventory management practices, these four hospitals are considered as one entity. At the beginning of the day, hospitals receive platelet products that were ordered on the previous day, from CBS. In the case of shortages, hospitals can place expedited (same-day) orders at a higher cost. Prior to September 2017, platelets had five days of shelf life, while after this date, the shelf life of platelets was increased to seven days. After exceeding the shelf life, platelet products are expired and discarded.
3.2 Data description

The data in this study are constructed by processing CBS shipping data and the TRUST (Transfusion Research for Utilization, Surveillance and Tracking) database at the McMaster Centre for Transfusion Research (MCTR) for platelet transfusion in Hamilton hospitals from 2010/01/01 to 2018/12/31. The study is approved by the Canadian Blood Services Research Ethics Board and the Hamilton Integrated Research Ethics Board (HiREB number 7293). The need for consent was waived by the ethics committee, and the data were de-identified to ensure that individual patients cannot be directly identified by the authors. The data are high dimensional, with more than 100 variables that can be divided into four main groups: 1. the blood inventory data such as product name and type, received date, expiry date, 2. patient characteristics such as age, gender, patient ABO Rh blood type, 3. the transfusion location such as intensive care, cardiovascular surgery, hematology, and 4. available laboratory test results for each patient such as platelet count, hemoglobin level, creatinine level, and red cell distribution width. The laboratory tests are prescribed by physicians based on clinical needs and can help to decide whether a patient needs platelet transfusion. In this research, the laboratory test results are processed and used along with other information to forecast future platelet demand.

Additionally, we add new calculated predictors such as the number of platelet transfusions in the previous day and previous week, the number of received units in the previous day, and day of the week. Table 1 gives the set of predictors that are used in this study along with their descriptions. These predictors are selected by a lasso regression model [26] which is explained in detail in Section 3.3.2. As we can see from Table 1, predictors have different ranges, and hence are standardized by z-score normalization. All data processing and analysis and model implementations are carried out using the Python 3.7 programming language.

3.3 Demand forecasting models

This section explains the five forecasting models used for forecasting the platelet demand in Hamilton hospitals. The ARIMA and Prophet models are univariate models that forecast the demand based on demand history. Lasso regression, random forest, and LSTM networks are multivariable models that consider various predictors in addition to demand history for forecasting the demand.

3.3.1 Univariate models. The univariate models, ARIMA and Prophet, forecast the demand solely based on the previous demand values. The ARIMA model does not consider seasonality in data and is considered as a baseline model. The Prophet model, on the other hand, considers trend, seasonality, and holidays for forecasting the demand.

**ARIMA model.** An autoregressive integrated moving average model consists of three components, an autoregressive (AR) component that considers a linear combination of lagged values as the predictors, a moving average (MA) component of past forecast errors (white noise), and an integrated component where differencing is applied on the data to make it stationary. Let \( y_1, y_2, \ldots, y_t \) be the demand values over time period \( t \); the time series data can be written as:

\[
y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, y_{t-n}) + \epsilon_t
\]

An ARIMA model assumes that the value of demand is a linear function of a number of previous past demand values and previous error values. Thus, the ARIMA model can be written as:

\[
\hat{y}_t = \mu + \vartheta_1 y_{t-1} + \vartheta_2 y_{t-2} + \vartheta_3 y_{t-3} + \cdots + \vartheta_p y_{t-p} + \epsilon_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2} - \phi_3 \epsilon_{t-3} - \cdots - \phi_q \epsilon_{t-q}
\]
where \( y_t \) is the response variable (the predicted demand), \( \mu \) is a constant, \( \theta_i \) and \( \phi_j \) are model parameters in which \( i = 1, 2, \ldots, p \) and \( j = 0, 1, 2, \ldots, q \), \( p \) and \( q \) are the model orders and define the number of autoregressive terms and moving average terms, respectively.

ARIMA is used as the baseline model for comparison. In order to fit an ARIMA model, first the ADF test is applied on the time series data to examine the stationarity, and the standard auto_arima() function in Python is used for hyperparameter tuning and determining the optimal model order. A function is developed in Python to implement the ARIMA model via a rolling-origin strategy.

Prophet model. Prophet is a time series model introduced by [27] that considers common features of business time series: trends, seasonality, holiday effects and outliers. The Prophet model was developed for forecasting events created on Facebook and is implemented as an open source software package in both Python and R. Its scalability makes it well-suited for large datasets, and its ease of use and interpretation makes it accessible to a wide range of users. Let \( g_t \) be the time series trend function which shows the long-term pattern of data, \( s_t \) be the seasonality which captures the periodic fluctuations in data such as weekly, monthly or yearly patterns, and finally \( h_t \) be the non-periodic holiday effect. These features are combined

| Name             | Description                                           |
|------------------|-------------------------------------------------------|
| abnormal_ALP     | Number of patients with abnormal alkaline phosphatase |
| abnormal_MPV     | Number of patients with abnormal mean platelet volume |
| abnormal_hematocrit | Number of patients with abnormal hematocrit          |
| abnormal_PO2     | Number of patients with abnormal partial pressure of oxygen |
| abnormal_creatinine | Number of patients with abnormal creatinine          |
| abnormal_INR     | Number of patients with abnormal international normalized ratio |
| abnormal_MCHb    | Number of patients with abnormal mean corpuscular hemoglobin |
| abnormal_MCHb_conc | Number of patients with abnormal mean corpuscular hemoglobin concentration |
| abnormal_hb      | Number of patients with abnormal hemoglobin          |
| abnormal_mcv     | Number of patients with abnormal mean corpuscular volume |
| abnormal_plt     | Number of patients with abnormal platelet count       |
| abnormal_redcellwidth | Number of patients with abnormal red cell distribution width |
| abnormal_wbc      | Number of patients with abnormal white cell count     |
| abnormal_ALC     | Number of patients with abnormal absolute lymphocyte count |
| location_GeneralMedicine | Number of patients in general medicine            |
| location_Hematology | Number of patients in hematology                      |
| location_IntensiveCare | Number of patients in intensive care                |
| location_CardiovascularSurgery | Number of patients in cardiovascular surgery |
| location_Pediatric          | Number of patients in pediatrics                     |
| Monday            | Indicating the day of the week                        |
| Tuesday           | Indicating the day of the week                        |
| Wednesday         | Indicating the day of the week                        |
| Thursday          | Indicating the day of the week                        |
| Friday            | Indicating the day of the week                        |
| Saturday          | Indicating the day of the week                        |
| Sunday            | Indicating the day of the week                        |
| lastWeek_Usage    | Number of units transfused in the previous week       |
| yesterday_Usage   | Number of platelet units transfused in the previous day |
| yesterday_ReceivedUnits | Number of units received by the hospital in the previous day |

https://doi.org/10.1371/journal.pone.0297391.t001
through a generalized additive model (GAM) [28], and the Prophet time series model can be written as:

$$\hat{y}_t = g_t + s_t + h_t + \epsilon_t$$  \hspace{1cm} (3)

The normally distributed error $\epsilon_t$ is added to model the residuals. We use the Prophet library in Python for implementing the Prophet model and develop a function for implementation via a rolling-origin strategy.

### 3.3.2 Multivariable models

In order to explore the effect of including clinical predictors in the forecasting process, in the next step we introduce three multivariable models that incorporate clinical predictors: lasso regression, random forest, and LSTM networks. These machine learning models are implemented to forecast the demand based on demand history and multiple predictors. Lasso regression is used as a forecasting model and a variable selection method to select the most relevant predictors for the multivariable models.

**Lasso regression.** We use lasso regression [26] since it allows predictors to be included in the demand forecasting process. Lasso uses an L1 penalty, which tends to push some coefficients towards exactly zero, hence it performs variable selection by reducing the impact of irrelevant or less important predictors. This leads to a reduction in model complexity while improving the prediction accuracy. By considering the actual demand on day $t$ ($t = 1, 2, \ldots, N$) as $y_t$ and the predicted demand on day $t$ as the product of the clinical predictors ($z_{ij}$) and their corresponding coefficients $\beta_j$, where $j = 1, 2, \ldots, M$ specifies the clinical predictor, the lasso regression is the solution to the following optimization problem:

$$\sum_{t=1}^{N} (y_t - \sum_j \beta_j z_{ij})^2 + \lambda \sum_{j=1}^{M} |\beta_j|$$  \hspace{1cm} (4)

subject to $\sum_{j=1}^{M} |\beta_j| \leq \tau$.  \hspace{1cm} (5)

The optimization problem defined in (4)-(5) chooses the coefficients, $\beta$, that minimize the sum of squares of the errors between the actual values ($y$) and the response variable values, with a sparsity penalty ($\lambda$) on the sum of the absolute values of the model coefficients. Constraint (5) forces some of the coefficients (that have a minor contribution to the estimate) to be zero. Predictors that have non-zero coefficients are selected in the model, and the response variable is calculated as follows:

$$\hat{y}_t = \beta z_t$$  \hspace{1cm} (6)

In this study, lasso regression is used as a variable selection method to find important predictors for platelet demand. Subsequently, this information is used for demand forecasting. We use the LassoCV function from the sklearn package in Python to implement the lasso regression. The penalty coefficient $\lambda$ is chosen through five-fold cross-validation. A function is developed to implement the lasso regression via a rolling-origin strategy.

**Random forest.** Random forests, first proposed in [29], are ensemble methods that use decision trees. We chose to explore random forests as they can capture nonlinear relationships between predictors while also being interpretable, as what a decision tree does can be understood by simply looking at it. Decision trees in a forest are trained using bootstrapped samples and are only allowed to consider a subset of the predictors when choosing splits. Considering the actual demand on day $t$ as $y_t$, and the set of days in the bootstrap samples as $D$, a tree starts
with a root node that has an attached value $\mu$:

$$\mu = \frac{1}{|D|} \sum_{i \in D} y_i$$  \hspace{1cm} (7)$$

This node creates two child nodes that separate data points based on a clinical predictor, $u$, where one node gets data with the value of $u$ on day $t$ ($z_{tu}$) less than a value $v$ and the other node gets data with $z_{tu}$ greater than or equal to $v$. These child nodes have attached values calculated in the same way as the root, $\mu_1 = \frac{1}{|[z_{tu} < v]|} \sum_{t \in D, z_{tu} < v} y_i$ and $\mu_2 = \frac{1}{|[z_{tu} \geq v]|} \sum_{t \in D, z_{tu} \geq v} y_i$.

The split measures, $u$ and $v$, are chosen by minimizing the variance of the model. A random forest grows a number of these trees, $K$, and produces a prediction for a set of clinical predictors, $z_n$ by averaging together the predictions of each of the trees:

$$\hat{y}_t = \sum_{i=1}^{K} T_i(z_t)$$  \hspace{1cm} (8)$$

where each tree $T_i$ takes a set of clinical predictors and traverses the nodes of tree $i$ using the splits found as described above. Forecasting problems can have linear or nonlinear relationships among the model predictors. Random forests can work on both linear and nonlinear data, and are able to capture nonlinear dependencies among the predictors. We use the RandomForestRegressor function from the scikit-learn package in Python to implement the random forest. Hyperparameter tuning is achieved by using grid search on the number of trees, maximum tree depth, and the number of features to consider when looking for the best split. The best split in a tree is chosen by minimizing MSE (Mean Square Error) and five-fold cross-validation is used to reduce overfitting. We developed a function in Python to implement the random forest model via a rolling-origin strategy.

**LSTM network.** LSTM networks are a class of recurrent neural networks (RNN) that were introduced by [30] and are capable of learning long-term dependencies in sequential data. In theory, RNNs should be capable of learning long-term dependencies, however they suffer from the so-called vanishing gradient problem. Consequently, LSTM networks are designed to resolve this issue. An LSTM network maps a set of input neurons (also called units) to a set of output neurons through a hidden layer. A neuron or unit in an LSTM network consists of an input gate ($i_t$), a forget gate ($f_t$), a cell state ($c_t$), and an output gate ($o_t$). For a more comprehensive explanation of how LSTM networks are implemented, we refer the reader to [31].

The hidden layer output can be written as a function of the gates, the model input (here the clinical predictors ($z_t$)), and the previous output of the hidden layer:

$$h_t = \sigma_h(i_t, f_t, c_t, o_t, z_t, h_{t-1})$$  \hspace{1cm} (9)$$

The output of the LSTM network, here the demand forecasts, is calculated as a weighted sum of the hidden layer outputs plus a bias, $b$:

$$\hat{y}_t = wh_t + b$$  \hspace{1cm} (10)$$

Like random forests, LSTM networks are able to capture nonlinear dependencies among the predictors. We implement the LSTM network using the TensorFlow package [32]. The LSTM network is trained by using the ADAM optimizer [33], and MSE is used as the loss function for this optimizer. For hyperparameter tuning, grid search is performed to find the best model parameters (including the number of epochs, batch size, and number of hidden layers) toward the minimum MSE. Moreover, 10-fold cross-validation is used to reduce overfitting. A
function is developed in Python to implement the LSTM network model via a rolling-origin strategy.

3.4 Empirical evaluation

3.4.1 Exploratory analysis for trends, seasonality and holiday patterns. In order to propose a short-term demand forecasting model, we first explore the data for identifying temporal (daily/monthly) patterns that can inform our demand forecasting techniques. In particular, we investigate correlations among the predictors, seasonality, day of the week, and non-stationarity effects.

Identify non-stationarity. The Augmented Dickey-Fuller (ADF) test [34] is applied on the time series data to examine the stationarity.

Identify seasonality. We apply the Seasonal and Trend decomposition using Loess (STL) model to decompose the time series data into trend, seasonality, and residuals. We also apply the Kruskal-Wallis test to compare the transfused units in different months, and apply the Mann-Whitney test for comparing the transfused units on weekdays vs. weekends. Moreover, we explore the trend, holidays, weekly seasonality, and yearly seasonality using the Prophet model, explained in Section 3.3.1.

Identify day of the week effect. We also compare the mean daily units transfused based on day of the week by plotting the mean against day of the week, and also by applying the t-test to compare the mean daily units transfused during weekdays and weekends.

3.4.2 Rolling window analysis. We fit the forecasting models multiple times in order to collect multiple out-of-sample one-step ahead forecast errors by using a rolling window. The rolling window is used as part of the demand forecasting process to periodically retrain the models and use more recent data. The flowchart of the proposed demand forecasting system is given in Fig 2. We retrain each model periodically, according to two parameters, the training window and the retraining period. When we retrain a model, we use a training window of the most recent data. For evaluation, we consider a rolling-origin evaluation, similar to the one presented in [35]. Many studies consider a fixed-origin evaluation, but we consider a rolling-origin evaluation to improve the efficiency and reliability of out-of-sample tests [35]. In a rolling-origin evaluation, the forecasting origin is successively updated and new forecasts are produced from each new origin. We set the forecasting window and rolling steps to be the same as the retraining period.

Here we consider two training windows, two years (starting from 2016) and eight years (starting from 2010), to explore the impact of data volume. The forecasting horizon is one year (2018) in which next day forecasts are generated for each of the retraining periods. We consider retraining periods of 1, 7, 30, and 90 days, to examine the trade-off between the accuracy and the overhead of retraining. The forecasting accuracy is computed by averaging the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE) over the forecasting horizon for each rolling origin.

4 Results

This section presents the results of the exploratory analysis for trends, seasonality and holiday patterns. We also present demand forecasting comparisons for univariate and multivariable models, and the forecasting performance of the models trained with training window sizes of two and eight years and retraining periods of 1, 7, 30, and 90 days. We implement the models to forecast the daily demand aggregated over four hospitals for one day ahead via a rolling-
origin strategy. We periodically retrain our models based on the rolling window analysis explained in Section 3.4.2.

### 4.1 Trends, seasonality and holiday patterns

The data analysis ranges from 2010/01/01 to 2018/12/31. An initial observation is that the demand is highly variable, with a transfused daily average of 17.90 units and a standard deviation of 7.05 units.

**Observations for non-stationarity.** The results of the ADF test show that the data is not stationary ($P$ value = 0.085) before 2016, but it becomes stationary from 2016 onwards ($P$ value < 0.001).

**Observations for seasonality.** Fig 3 shows the time series data decomposition using the STL model. As we can see in the seasonal part, there are recurring temporal patterns in the data. The results of the Mann-Whitney U test also show that there is a significant difference between the distributions of daily transfusions during weekdays and weekends (weekday = mean [sd]: 21.20 [6.22], weekend = mean [sd]: 12.37 [4.60], Mann-Whitney U test: ...)
The results of the Kruskal-Wallis test show that there is a considerable difference in the distributions of daily transfusions in different months ($\chi^2 = 39.28$, $P$ value $< 0.001$), which provide strong evidence in favour of the presence of weekly and monthly seasonalties. Since the data becomes stationary from 2016 onwards, we also explore the trend, holidays, weekly seasonality, and yearly seasonality (seasonality within a year) starting from 2016 using the Prophet model. As we can see from Fig 4, there is a downward trend from the beginning of 2016 to July 2017 and an upward trend from July 2017 to the end of 2018. Almost all holidays have a negative effect on the model, except for July 1st. This means that the demand is lower than regular weekdays for almost all of the holidays, except for July 1st.

We can also see that there is weekly seasonality in which Wednesdays have the highest demand while the weekends have the lowest demand. Moreover, the yearly seasonality, captured by the Fourier series in the Prophet model, depicts three cycles: 1. January to May in which March has the highest demand while May has the lowest demand; 2. May to September in which the demand is highly variable. July has the highest demand in this cycle and the highest demand of all months while May has the lowest demand in the cycle and also the lowest demand of all months; 3. September to January with a slight variation in demand—November with the highest and January with the lowest demands.
Observations for day of the week effect. Fig 5 compares the mean daily units transfused based on day of the week, month, and year. As we can see from Fig 5C, there is a significant difference in the mean daily platelet usage when comparing weekdays to weekends (weekday = mean [sd]: 21.20 [6.22], weekend = mean [sd]: 12.37 [4.60], t-test: 95% confidence interval for the difference in means: (7.97, 10.34), \( P \) value < 0.001). Consequently, there is a clear weekday/weekend effect, in agreement with Fig 4, which appears to be caused by various reasons including lower staffing levels and operating hours over the weekends and
prophylactic platelet transfusions to cancer patients on Fridays to ensure that their platelet counts remain sufficiently high over the weekend.

### 4.2 Demand forecasting comparisons for univariate models

Fig 6 compares the forecasts generated by the univariate models, with a training window of two years and by retraining every day, and the actual demand. The actual demand has a large variance (mean [sd]: 19.28 [7.36]). The ARIMA model’s forecasts have significantly lower variance (mean [sd]: 18.89 [3.09]) in comparison to the actual demand, meaning that the forecasts cannot capture the wide range of the actual demand. Despite having a larger variance than the ARIMA model, Prophet shows a similar behavior (mean [sd]: 19.35 [4.40]).

Next, we examine the univariate models’ residuals via the ACF (Autocorrelation Function). Fig 7 gives the coefficients of correlation between a value and its lag for ARIMA and Prophet.
As we can see in Fig 7A, there is an autocorrelation at time seven (and multiples of seven) due to weekly seasonality that is not incorporated in the model. Since seasonality is one of the primary features of our time series data, we include seasonality directly in the forecasting process by using the Prophet model. As we can see in Fig 7B, there is no repeated autocorrelation pattern for Prophet residuals.

We also perform a pairwise t-test to compare the univariate models’ residuals. The results show a statistically significant difference between the ARIMA residuals (mean [sd]: 0.39 [6.80]) and Prophet residuals (mean [sd]: -0.07 [5.90], t-test: 95% confidence interval for the difference in means: (0.08, 0.85), P value = 0.018), indicating higher residuals in the ARIMA model.

4.3 Demand forecasting comparisons for multivariable models

We begin this section with an examination of selecting the clinical predictors for the multivariable models. Next, we compare the forecasts generated by the multivariable models and the actual demand.

4.3.1 Selecting the predictors using Lasso regression. As discussed in Section 3.2, the data has more than 100 features, and we select predictors via lasso regression. The 29 clinical predictors that are introduced in Section 3.2 are selected by lasso regression and used for training the multivariable models. One of the data characteristics is that clinical predictors are highly correlated. These high correlations can affect the performance of a regression model, mainly because of the violation of model assumptions.

We calculate the Pearson correlation between the selected predictors. The Pearson correlation measures the linear relationship between two variables, ranging from -1 to 1, where -1 corresponds to a perfect negative correlation and 1 corresponds to a perfect positive correlation. As shown in Fig 8, the predictors, in particular the daily numbers of patients with abnormal laboratory test results, are highly correlated. These high correlations give rise to some challenges when the predictors are considered in the demand forecasting process, as discussed in Table S1 of S1 Appendix.

We also calculate the confidence intervals for these clinical predictors (also referred to as the model predictors). There are multiple methods for calculating a confidence interval for the predictors; one of the most popular is the bootstrap method [36]. The bootstrap method is used in the experiments for calculating the confidence intervals for the predictors used in the multivariable models. As shown in Fig 9, the predictors’ coefficients have a wide range, so we see high values (abnormal_plt = 0.23) as well as low values (Friday = -0.39) for the lab tests and day of the week. Overall, we can see that the range of the predictors’ coefficients for the 95% confidence interval is narrow. Detailed information about the predictors and their corresponding coefficients are given in Table S1 of S1 Appendix.

4.3.2 Comparisons of multivariable models forecasts. Fig 10 shows the actual daily units transfused and the forecasts generated by the multivariable models, lasso regression, random forest and LSTM network, with a training window of two years and by retraining every day. The forecast means of lasso regression (mean [sd]: 19.12 [3.62]) and random forest (mean [sd]: 19.72 [4.28]) are very close to the actual mean demand, but forecast standard deviations are much lower than the actual demand standard variation. LSTM network forecasts have a slightly lower mean (mean [sd]: 18.01 [3.55]) but significantly lower standard deviation than the actual demand. Next, a repeated measures ANOVA test is performed for comparing the multivariable models’ residuals. The results of the test show a statistically significant difference between the lasso regression, random forest, and LSTM network residuals ($F = 35.86, P$ value <0.001). To show which models’ residuals are significantly different, we perform pairwise
comparisons by using a pairwise t-test. Table 2 gives the results of the pairwise t-test for the models’ residuals, showing that they are significantly different from each other. The P values are adjusted using the Bonferroni multiple testing correction method.

4.4 Performance comparisons

The performance of the forecasting models is computed based on a rolling-origin evaluation and by four error measures, RMSE, MAE, MAPE, and SMAPE. The first two error measures, RMSE and MAE, are absolute measures while the remaining ones, MAPE and SMAPE, are relative measures. The errors are measured for each rolling origin for the test data and reported in Figs 11–14 and Table 3. Table 3 gives the mean and standard deviation of the errors for different training window sizes and retraining periods.

Fig 8. Pearson correlation among variables.

https://doi.org/10.1371/journal.pone.0297391.g008
Table 2. Comparison of multivariable models residuals using a pairwise t-test.

| Model         | Descriptive statistics | T-test                      |
|---------------|------------------------|-----------------------------|
|               | Mean | Standard Deviation | Model | 95% confidence interval for the difference in means | P value |
| Lasso Regression | 0.16 | 6.39 | Random Forest | (0.09, 1.12) | 0.020 |
| Random Forest | -0.44 | 8.77 | LSTM Network | (1.60, 1.83) | <0.001 |
| LSTM Network | 1.27 | 8.34 | Lasso Regression | (-1.57, -0.65) | <0.001 |
the size of the training window mostly affects the univariate models, ARIMA and Prophet. ARIMA’s performance improves when moving from two years to eight years of data. Since ARIMA’s forecasts are only based on the previous demands, and the seasonality in data has not changed significantly during the eight years, the model parameters, $p$ and $q$, are more robust for longer time series data (including 5 lagged values and a moving average of 2).

Fig 11. RMSE with different training window sizes and retraining periods.

https://doi.org/10.1371/journal.pone.0297391.g011
resulting in more accurate forecasts. In general, when a limited amount of data are available, the ARIMA model has a high forecast error not only because its forecasts are solely based on the previous demands, but also due to the fact that it cannot capture the seasonality in the data. Prophet’s accuracy is also improved as the amount of data increases. However, unlike ARIMA, the forecast errors are similar for different retraining periods. The results for the lasso

Fig 12. MAE with different training window sizes and retraining periods.

https://doi.org/10.1371/journal.pone.0297391.g012
regression and LSTM network indicate that there is not much difference for these methods when there is a large amount of data for training, or when different retraining periods are considered. Random forest does see a slight improvement with eight years of data, and it is the only multivariable model to see this improvement. Its forecast errors are very close for different retraining periods.

Fig 13. MAPE with different training window sizes and retraining periods.

https://doi.org/10.1371/journal.pone.0297391.g013
In terms of the retraining periods, retraining the models less frequently reduces the variability of the error. If we compare Fig 11A with Fig 11G, we see that the RMSE error is less variable in Fig 11G for all the models, similarly for MAE in Fig 12. This can also be verified from the results in Table 3, where we see lower standard deviations as we move down to retraining every 90 days.

Fig 14. SMAPE with different training window sizes and retraining periods.

https://doi.org/10.1371/journal.pone.0297391.g014

In terms of the retraining periods, retraining the models less frequently reduces the variability of the error. If we compare Fig 11A with Fig 11G, we see that the RMSE error is less variable in Fig 11G for all the models, similarly for MAE in Fig 12. This can also be verified from the results in Table 3, where we see lower standard deviations as we move down to retraining every 90 days.
Table 3. Model performance with different training window sizes and retraining periods.

| Retraining Period | Model         | RMSE  | MAE   | MAPE  | SMAPE | RMSE  | MAE   | MAPE  | SMAPE |
|-------------------|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 day             | ARIMA         | 7.44  | 7.72  | 19.46 | 12.55 | 21.37 | 14.01 | 5.72  | 23.8   |
|                   | Prophet       | 4.20  | 3.19  | 18.52 | 12.85 | 19.05 | 13.12 | 4.28  | 12.3   |
|                   | Lasso Regression | 4.88  | 4.40  | 19.66 | 12.86 | 18.92 | 12.25 | 4.80  | 13.39  |
|                   | Random Forest | 4.59  | 4.32  | 19.49 | 12.58 | 20.23 | 13.05 | 4.59  | 13.27  |
|                   | LSTM Networks | 4.31  | 3.99  | 19.41 | 12.69 | 19.23 | 12.36 | 3.09  | 2.49   |
| 7 days            | ARIMA         | 6.81  | 5.1    | 33.19 | 28.33 | 30.02 | 7.74  | 5.85  | 4.58   |
|                   | Prophet       | 7.34  | 5.01  | 32.32 | 8.36  | 28.65 | 8.25  | 5.62  | 1.69   |
|                   | Lasso Regression | 6.22  | 5.15  | 31.77 | 7.50  | 27.74 | 4.94  | 6.31  | 1.80   |
|                   | Random Forest | 6.43  | 5.17  | 27.48 | 8.70  | 25.64 | 6.62  | 6.12  | 1.94   |
|                   | LSTM Networks | 4.32  | 2.01  | 26.45 | 8.39  | 25.83 | 7.25  | 4.07  | 1.82   |
| 30 days           | ARIMA         | 7.18  | 1.22  | 32.49 | 5.34  | 30.46 | 3.43  | 4.86  | 0.94   |
|                   | Prophet       | 6.62  | 1.23  | 36.92 | 5.06  | 32.38 | 8.40  | 4.85  | 1.20   |
|                   | Lasso Regression | 5.86  | 0.74  | 30.25 | 5.14  | 25.79 | 3.48  | 5.44  | 0.89   |
|                   | Random Forest | 6.13  | 1.61  | 29.09 | 6.41  | 25.93 | 4.33  | 5.41  | 1.36   |
|                   | LSTM Networks | 5.41  | 0.81  | 28.06 | 5.61  | 24.68 | 3.98  | 4.99  | 1.18   |
| 90 days           | ARIMA         | 7.44  | 0.50  | 35.68 | 3.10  | 31.29 | 2.36  | 5.08  | 0.72   |
|                   | Prophet       | 6.23  | 0.45  | 35.13 | 4.51  | 32.27 | 3.07  | 4.95  | 0.39   |
|                   | Lasso Regression | 6.38  | 0.50  | 32.71 | 5.83  | 24.82 | 1.81  | 4.94  | 0.58   |
|                   | Random Forest | 6.06  | 1.27  | 28.89 | 3.53  | 25.59 | 1.92  | 5.20  | 1.85   |
|                   | LSTM Networks | 5.26  | 0.63  | 27.76 | 3.74  | 24.29 | 1.64  | 5.48  | 0.55   |

Figs 13 and 14 compare the MAPE and SMAPE of the models trained with different training window sizes and retraining periods. As we can see from these figures and from Table 3, increasing the training window size does not necessarily decrease the errors. There is similar behavior for the retraining periods, especially for the multivariable models, but we see that retraining less frequently results in less variable errors.

Overall, the results indicate that while univariate models can benefit from a larger training window size and frequent retraining, the performance of the multivariable models is not affected by a larger training window, meaning that these models have robust performance with different data volumes.

5 Comparison and discussion

In this section, we compare the models and provide recommendations for using these models in various scenarios. In Section 5.1 we compare the models based on a training window of two years, in Section 5.2 we discuss the impact of an increased amount of data on the forecasting models, and in Section 5.3 we discuss the effect of different retraining periods on the models. Finally, in Section 5.4 we provide the overall methodological implications of the study and in Section 5.5 discuss managerial implications.

5.1 Univariate versus multivariable models

We have presented five different models for platelet demand forecasting that can be divided into two groups: univariate and multivariable. Univariate models, ARIMA and Prophet, forecast future demand based only on the demand history. Although the ARIMA model only considers a limited number of previous values for forecasting the demand, retraining it every day,
week or month leads to a slight performance improvement. The Prophet model incorporates the historical data, seasonality and holiday effects into the demand forecasting model which results in an improvement in the forecasting accuracy by approximately 10% compared to ARIMA. This highlights the impact of weekday/weekend and holiday effects in the platelet demand variation. As we discussed in Section 4.1, there is a weekday/weekend effect for platelet demand, which is not (directly) captured in the ARIMA model.

Multivariable models, on the other hand, incorporate clinical predictors as well as historical demand data for demand forecasting. We use lasso regression to select the dominant clinical predictors that affect the demand. Lasso regression examines the linear relationship among the clinical predictors and their influence on the demand. However, as presented in Fig 8, there are correlations among the clinical predictors. There may also be nonlinear relationships among these clinical predictors that cannot be captured by a linear regression model. These issues motivated us to use two machine learning approaches, random forest and LSTM network. Random forest is capable of capturing nonlinearities among variables and its forecasting method of averaging past values provides some contrast to the LSTM network’s modelling approach. An LSTM network can also account for nonlinearities among variables. Moreover, an LSTM network is capable of retaining past information while forgetting some parts of the historical data. As we can see from Table 3, in general, random forest, LSTM network and lasso regression have low forecast errors for different training window sizes, owing to the inclusion of the clinical predictors.

5.2 Two years versus eight years of data
As discussed in Section 3.3, we train our models for two training window sizes, with two years and eight years of data, respectively. Since there is no trend in the data from 2016 onwards (see Fig 3), in the first scenario the models are trained for two years (training window size of two years, starting from 2016). With this amount of data and by retraining every year, forecasts are not accurate for univariate time series approaches, and one needs to include the clinical predictors in the forecasting model. However, by considering a training window of eight years, the ARIMA model’s performance improves by approximately 20%, compared to the case of a two year training window. The Prophet model’s performance also improves when more data are available, specifically when it is trained less frequently (30 and 90 day retraining windows).

In general the multivariable models result in small forecasting errors for two years of data for training, and do not perform significantly better as the amount of data increases, which shows that there is not much sensitivity to the training window size. This highlights the importance of including the clinical predictors in the forecasting process.

5.3 Different retraining periods
We also compare different retraining periods and provide insight on how to choose the appropriate retraining period for this data (and in general). Our results show that considering different retraining periods does not affect the models in the same manner. While in general all the models benefit from retraining more frequently, univariate models benefit more. For the univariate models, the greatest performance increase is for the ARIMA model when retrained every day, resulting in a decrease of 50% in MAPE and SMAPE. For the multivariable models, lasso regression has an impressive performance increase when retrained every day, while random forest and LSTM networks show less sensitivity to the retraining period. So, by considering the overhead of retraining these models more frequently, one may decide to use a larger retraining window for random forest and LSTM networks.
Generally, if the retraining period is small, meaning that the models are retrained more frequently, the mean forecast accuracy representing the long-term overall performance is improved.

5.4 Methodological implications of the study

In general, when there is access only to previous demand values, using a univariate model and retraining it frequently is effective. In practice, applying forecasting models to real-world healthcare systems can be challenging. One common concern is data accessibility, especially in small or rural hospitals. In such healthcare facilities, laboratory test results may not be available or may be limited. In these systems, we recommend employing a simple model such as a univariate time series that forecast the demand based on the historical demand of platelets. Our results suggest that using a simple univariate time series model can yield results comparable to those of more complicated models, in particular when sufficient data is available. Moreover, in resource-limited settings where access to data scientists and statisticians is limited, a straightforward univariate model like ARIMA or Prophet can be employed. These models do not demand extensive expertise for training and usage, making them more accessible and suitable for such environments. Notably, the World Health Organization (WHO) suggests that alongside the requirement for robust blood forecasting models, there is also a growing need for simple models that can be employed in all settings. Univariate models align well with these requirements [37].

In the case that several data variables are available, lasso regression, random forest models and LSTM networks can forecast the demand with higher accuracy even when a small amount of data is available and without frequent retraining. Forecasting problems can have linear or nonlinear relationships among the model variables. Due to the fact that LSTM networks are appropriate for both linear and nonlinear time series, and are able to capture nonlinear dependencies, they can outperform linear regression models when long term correlations exist in the time series. Based on the LSTM results, we conclude that long term correlations and nonlinearity are not major issues for our data since the LSTM model does not significantly outperform lasso regression.

While LSTM networks perform well even with a limited amount of data and they can capture nonlinear relationships, they lack interpretability. Interpretability is an important feature of any prediction model used in a safety critical setting like blood product distribution. Considering the time and memory complexity, and interpretability of these models, lasso regression has lower time and memory complexity while it is also very interpretable. Random forest models maintain interpretability while also having the ability to capture nonlinear relationships. Random forests do well when their training data has good coverage of the different feature combinations the model is forecasting. This is because random forest models make forecasts for a set of features by averaging together similar data points from the training data. This allows random forest models to extract nonlinear relationships but also means they cannot extract trends effectively and may need a large amount of data in order to work well. This can be seen in our model (see Table 3), a training window of eight years, with more training data points to reference, has a small improvement in the error measures over a training window of two years for different retraining periods.

Training random forest models and LSTM networks requires expertise in the machine learning area since poor training will cause low-precision results. It is also worth mentioning that the LSTM network is a robust learning model and is capable of learning linear and nonlinear relationships among the model variables even in very short time series data [38, 39]. However, as the number of inputs increases, both the data variables that make data wide and the
data rows that make data tall, LSTM performance tends to decrease because it is highly dependent on the input size. Moreover, wide data results in model overfitting [40]. Having wide data, one can apply a feature selection method such as lasso regression to reduce the number of variables and regularize the input.

One limitation of forecasting models is that they cannot capture sharp peaks in demand. Fig 15 depicts the actual and predicted demands for the second half of 2018 with a training window of two years and a retraining period of 7 days (retraining weekly) using lasso regression. It appears that the model does some degree of smoothing and thus cannot detect the sharp peaks. One possible explanation is that regression models are regressed on the expectation of the outcome, and are not good at capturing the extreme deviations from this expectation. However, as shown in Fig 15, smoothing mostly occurs for the maxima rather than the minima. In other words, the model potentially has large errors when there is excess demand, for example in emergency situations. The results presented in Sections 4.2 and 4.3 support that all the models struggle with capturing the peaks in demand. To sum up, when a sufficient amount of data is available, using a univariate model results in a low forecast error, particularly in the case that it is retrained every day. Specifically, when there is only access to the previous demand (as is currently the case for CBS) and adequate historical data are available, one can benefit from a simple univariate model like ARIMA or Prophet, since univariate models are simpler than the multivariable models. Multivariable models are useful when there is access to a limited amount of data. Also, they do not necessarily require frequent retraining, which may be an important implementation concern.

5.5 Managerial implications of the study

The short shelf life of platelets results in wastages which not only incur large costs but also affect the environment since they cannot be reused, recycled, or recovered [41]. Moreover, since platelet demand is highly variable, urgent same-day deliveries are placed frequently. Apart from the high cost of urgent orders, platelet shortage can increase the risk of putting patients’ lives in danger. Currently, blood suppliers are not aware of the demand at the hospitals since hospitals hold excess inventory to manage the highly variable platelet demand.
Indeed, this has its roots in the bullwhip effect, that is hospitals tend to order more than their actual demand. We see an opportunity to better coordinate supply (number of units received) with demand (number of units transfused) through the development of a daily demand predictor. Forecasting the demand improves the transparency between blood suppliers and hospitals, and helps blood suppliers to make better-informed decisions.

From the clinical perspective, accurate demand forecasting is important for clinical and supply chain management purposes. Demand forecasting can be used for placing optimal platelet orders and for decision making in many parts of the supply chain such as donation planning, and resource and staff management. As we can see in Section 4, there appears to be a limit to how accurate the demand forecasts can be, so one important challenge would be how to use the demand forecasts to inform an ordering policy in an effective manner. Clearly, forecasts themselves do not reflect an optimal ordering decision but they can be used as additional information in building effective ordering/inventory management policies (incorporating such forecasts is one of our current research directions).

Moreover, this research provides a holistic analysis of the predictors that affect platelet demand, including the clinical predictors, hospital locations, day of the week and demand history. This can help blood suppliers with adapting clinically relevant factors into the decision making process, like decisions regarding the assignment of transfusion-related staff/resources (beds or equipment).

Overall, there is a significant caveat with all of these approaches in that there are still forecasting errors, in particular they all struggle with capturing peaks. These underestimations may cause significant concerns for using such forecasts directly as there is the danger of severe underestimation. Therefore, some adjustments may be required for using these forecasts according to specific objectives. For instance, one may need an optimization model for inventory control if the demand forecasts are used for inventory management.

6 Conclusion

In this study, we utilized two types of methods for platelet demand forecasting, univariate and multivariable methods. Univariate methods, ARIMA and Prophet, forecast platelet demand only by considering the historical demand information, while multivariable methods, lasso regression, random forest and LSTM networks, also consider clinical predictors. The error levels for the univariate models, particularly in the case that a small amount of data is available, motivates us to utilize clinical predictors to investigate their ability to improve the accuracy of forecasts. Results show that lasso regression, random forest and LSTM networks outperform the univariate methods when a limited amount of data are available. Moreover, since they include clinical predictors in the forecasting process, their results can aid in building a robust decision making and blood utilization system. However, their application is not limited to platelet products. We believe that they can be used in various areas, including healthcare in general, finance and climate studies, when data features are available. On the other hand, when there is access to a sufficient amount of data, the marginal improvement for a simple univariate model such as ARIMA is higher than for multivariable models. In such scenarios, univariate models can be applied to historical data for demand forecasting, regardless of the product, which makes these models generalizable and widely applicable.

Future extensions of this work will include: (i) proposing an optimal ordering policy based on the predicted demand over a planning horizon with ordering cost, wastage cost and shortage (same-day order) cost; (ii) further exploring the lasso regression approach to enhance variable selection, with a particular focus on interpretability (this will not only affect the lasso regression itself, but also may improve LSTM forecasting accuracy since LSTM inputs are
selected using lasso regression); (iii) more extensive empirical evaluation of the proposed models; (iv) exploring the generality of the results (outside of Hamilton).

Supporting information
S1 Appendix. Summary of most important predictors for platelet usage. (PDF)

Acknowledgments
The authors thank Tom Courtney, Rick Trifunov, Dr. John Blake, and Masoud Nasari for providing information about blood collection, blood processing, and blood distribution at Canadian Blood Services, and the anonymous reviewers of the manuscript for their comments that in particular improved the organization of the paper. All final decisions regarding manuscript content were made by the authors.

Author Contributions
Conceptualization: Maryam Motamedi, Na Li, Nancy M. Heddle.
Data curation: Maryam Motamedi, Na Li, Nancy M. Heddle.
Formal analysis: Maryam Motamedi, Douglas G. Down.
Funding acquisition: Na Li, Douglas G. Down.
Investigation: Jessica Dawson, Na Li.
Methodology: Maryam Motamedi, Jessica Dawson, Douglas G. Down.
Project administration: Maryam Motamedi, Na Li.
Resources: Douglas G. Down.
Software: Maryam Motamedi.
Supervision: Na Li, Douglas G. Down.
Validation: Maryam Motamedi, Na Li, Douglas G. Down.
Visualization: Maryam Motamedi, Jessica Dawson.
Writing – original draft: Maryam Motamedi, Jessica Dawson.
Writing – review & editing: Na Li, Douglas G. Down, Nancy M. Heddle.

References
1. Kumar A, Mhaskar R, Grossman BJ, Kaufman RM, Tobian AA, Kleinman S, et al. Platelet transfusion: a systematic review of the clinical evidence. Transfusion. 2015; 55(5):1116–1127. https://doi.org/10.1111/trf.12943 PMID: 25387589
2. Stanworth SJ, New HV, Apelseth TO, Brunskill S, Cardigan R, Doree C, et al. Effects of the COVID-19 pandemic on supply and use of blood for transfusion. The Lancet Haematology. 2020; 7(10):E756–E764. https://doi.org/10.1016/S2352-3026(20)30166-1 PMID: 32628911
3. Fontaine MJ, Chung YT, Rogers WM, Sussmann HD, Quach P, Galel SA, et al. Improving platelet supply chains through collaborations between blood centers and transfusion services. Transfusion. 2009; 49(10):2040–2047. https://doi.org/10.1111/j.1537-2995.2009.02236.x PMID: 19558430
4. Office of the Auditor General of Ontario. Value-for-Money Audit Blood Management and Safety; 2020. https://www.auditor.on.ca/en/content/annualreports/arreports/en20/20VFM_02bloodmgmt.pdf.
5. Cohen MA, Pierskalac WP. Target inventory levels for a hospital blood bank or a decentralized regional blood banking system. Transfusion. 1979; 19(4):444–454. https://doi.org/10.1046/j.1537-2995.1979.19479250182.x PMID: 473346

6. Hajijerna R. A new class of stock-level dependent ordering policies for perishables with a short maximum shelf life. International Journal of Production Economics. 2013; 143(2):434–439. https://doi.org/10.1016/j.ijpe.2011.05.021

7. Civelek I, Karaesmen I, Scheller-Wolf A. Blood platelet inventory management with protection levels. European Journal of Operational Research. 2015; 243(3):826–838. https://doi.org/10.1016/j.ejor.2015.01.023

8. Ensafian H, Yaghoubi S. Robust optimization model for integrated procurement, production and distribution in platelet supply chain. Transportation Research Part E: Logistics and Transportation Review. 2017; 103:32–55. https://doi.org/10.1016/j.ter.2017.04.005

9. Rajendran S, Ravindran AR. Inventory management of platelets along blood supply chain to minimize wastage and shortage. Computers & Industrial Engineering. 2019; 130:714–730. https://doi.org/10.1016/j.cie.2019.03.010

10. Guan L, Tian X, Gombars S, Zemek AJ, Krishnan G, Scott R, et al. Big data modeling to predict platelet usage and minimize waste in a tertiary care system. Proceedings of the National Academy of Sciences. 2017; 114(43):11368–11373. https://doi.org/10.1073/pnas.1714097114

11. Abouee-Mehrizi H, Mirjalli M, Sarhangian V. Data-driven platelet inventory management under uncertainty in the remaining shelf life of units. Production and Operations Management. 2022; 31(10):3914–3932. https://doi.org/10.1111/poms.13795

12. Elmachtoub AN, Grigas P. Smart “predict, then optimize”. Management Science. 2022; 68(1):9–26. https://doi.org/10.1287/mnsc.2020.3922

13. Li N, Chiang F, Down DG, Heddle NM. A decision integration strategy for short-term demand forecasting and ordering for red blood cell components. Operations Research for Healthcare. 2021; 29:100290. https://doi.org/10.1016/j.orhc.2021.100290

14. Critchfield GC, Connelly DP, Ziehmewin MS, Olesen CE, Scott EP. Automatic prediction of platelet utilization by time series analysis in a large tertiary care hospital. American Journal of Clinical Pathology. 1985; 84(5):627–631. https://doi.org/10.1093/ajcp/84.5.627 PMID: 4061386

15. Silva Filho OS, Cezarino W, Salviano GR. A Decision-making tool for demand forecasting of blood components. IFAC Proceedings Volumes. 2012; 45(6):1499–1504. https://doi.org/10.3182/20120523-3-RO-2023.00201

16. Silva Filho OS, Carvalho MA, Cezarino W, Silva R, Salviano G. Demand forecasting for blood components distribution of a blood supply chain. IFAC Proceedings Volumes. 2013; 46(24):565–571. https://doi.org/10.3182/20130911-3-BR-3021.00092

17. Kumari D, Wijayanyakake A. An efficient inventory model to reduce the wastage of blood in the national blood transfusion service. In: 2016 Manufacturing & Industrial Engineering Symposium (MIES). IEEE; 2016. p. 1–4.

18. Volken T, Buser A, Castelli D, Fontana S, Frey BM, Rüegs-Wolter I, et al. Red blood cell use in Switzerland: trends and demographic challenges. Blood Transfusion. 2018; 16(1):73–82. https://doi.org/10.2450/2016.0079-16 PMID: 27723455

19. Fanooodi B, Malmir B, Jahantigh FF. Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. Computers in Biology and Medicine. 2019; 113:103415. https://doi.org/10.1016/j.compbiomed.2019.103415 PMID: 31536834

20. Frankfurter GM, Kendall KE, Pegels CC. Management control of blood through a short-term supply-demand forecast system. Management Science. 1974; 21(4):444–452. https://doi.org/10.1287/mnsc.21.4.444

21. Fortsch SM, Khapalova EA. Reducing uncertainty in demand for blood. Operations Research for Health Care. 2016; 9:16–28. https://doi.org/10.1016/j.orhc.2016.02.002

22. Lebani F, Anwar U, Nuguah N, Azwar A. Forecasting demand in blood supply chain (case study on blood transfusion unit). In: Proceedings of the World Congress on Engineering, vol. 2; 2017.

23. Twumasi C, Twumasi J. Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of Tema General Hospital of Ghana. International Journal of Forecasting. 2022; 38(3):1258–1277. https://doi.org/10.1016/j.ijforecast.2021.10.008

24. Drackley A, Newbold KB, Paez A, Heddle N. Forecasting Ontario’s blood supply and demand. Transfusion. 2012; 52(2):366–374. https://doi.org/10.1111/j.1537-2995.2011.03280.x PMID: 21810099

25. Khalidi R, El Afia A, Chilbe R, Faizi R. Artificial neural network based approach for blood demand forecasting: Fez transfusion blood center case study. In: Proceedings of the 2nd International Conference on Big Data, Cloud and Applications; 2017. p. 1–6.
26. Tibshirani R. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996; 58(1):267–288.

27. Taylor SJ, Letham B. Forecasting at scale. The American Statistician. 2018; 72(1):37–45. https://doi.org/10.1080/00031305.2017.1380080

28. Hastie T, Tibshirani R. Generalized additive models: some applications. Journal of the American Statistical Association. 1987; 82(398):371–386. https://doi.org/10.1080/01621459.1987.10478440

29. Ho TK. Random Decision Forests. In: Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1)—Volume 1. ICDAR ’95. USA: IEEE Computer Society; 1995. p. 278.

30. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997; 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 PMID: 9377276

31. Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. Neural computation. 2000; 12(10):2451–2471. https://doi.org/10.1162/089976603762587262 PMID: 11032042

32. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for Large-Scale machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16); 2016. p. 265–283.

33. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014;

34. Cheung YW, Lai KS. Lag Order and Critical Values of the Augmented Dickey-Fuller Test. Journal of Business & Economic Statistics. 1995; 13(3):277–280. https://doi.org/10.1016/07350015.1995.10524601

35. Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting. 2000; 16(4):437–450. https://doi.org/10.1016/S0169-2070(00)00065-0

36. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. CRC press; 1994.

37.Organization WH, et al. WHO experts’ consultation on estimation of blood requirements: 03-05 February 2010, WHO-HQ, Geneva: meeting report. World Health Organization; 2010.

38. Boulmaiz T, Guermoui M, Boutagane H. Impact of training data size on the LSTM performances for rainfall–runoff modeling. Modeling Earth Systems and Environment. 2020; 6:2153–2164. https://doi.org/10.1007/s40808-020-00830-w

39. Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to diagnose with LSTM recurrent neural networks. arXiv preprint arXiv:151103677. 2015;

40. Lai G, Chang WC, Yang Y, Liu H. Modeling long-and short-term temporal patterns with deep neural networks. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval; 2018. p. 95–104.

41. Jemai J, Do Chung B, Sarkar B. Environmental effect for a complex green supply-chain management to control waste: A sustainable approach. Journal of cleaner production. 2020; 277:122919. https://doi.org/10.1016/j.jclepro.2020.122919