LOOK ACADEMIC PUBLISHERS
OPEN ACCESS

INTERNATIONAL JOURNAL OF ENVIRONMENTAL & SCIENCE EDUCATION
e-ISSN: 1306-3065
2018, Vol. 13, No. 6, 519-525

ARIMAX versus Holt Winter Methods: The Case of Blood Demand Prediction in Thailand

Ardthawee Tanyavutti [1*], Uthai Tanlamai [2]

[1] *Technopreneurship and Innovation Management Program (Chulalongkorn University), THAILAND*
[2] *Department of Accountancy (Chulalongkorn University), THAILAND*

* CORRESPONDENCE: ardthaweet@gmail.com

ABSTRACT

National and Regional Blood Center are the main blood transfusion centers to acquire blood from donors and distribute to hospitals in Thailand. However, the estimation of blood availability can be difficult because of specific factors such as a very high number of blood donors during the birthday of King and Queen and a seasonal blood shortage during major holidays like the New Year. This paper focuses on statistically prediction of blood demand in Thailand. Monthly data from January 2012 to December 2015 are separated into 2 periods, 45 months for model training and 3 months for model validation. Box-Jenkin's models with independent variables (ARIMAX) and Holt Winter's techniques are compared to report the best model fit using the smallest value of Mean Absolute Percentage Error. The two independent variables affecting the blood demands based on geolocation are Platelet Demand and Dengue Fever Patients. The result finds that majority predictions by ARIMAX provide better model fit. R script in Tableau is used for tool development.

Keywords: Auto Regressive Integrated with Moving Average and Exogenous Variables (ARIMAX), exponential smoothing, Red Blood Cell (RBC), National Blood Center (NBC), Mean Absolute Percentage Error (MAPE)

INTRODUCTION

Planning the future blood collection and usage requires adequate prediction of transfusion demand since blood is critical product to treat and save human lives. Blood is a perishable product which can be expired and is limited. Shortage of blood supply has been increasing for two decades, especially the most common blood product like the Red Blood Cell (RBC). National Blood Center (NBC) is the main blood transfusion center which obtains whole blood from donors. The center separates blood products and distributes them to nearby hospitals and Regional Blood Centers (Regions). However, the shelf life of donated blood is only 30 days which requires effective distribution planning (Owens, Tokessy, Rock, 2001). There is gap between blood demand and supply in Thailand at the current (**Figure 1**). The gap varies based on seasonal in the year. For example, there is a blood shortage during late December and early January of each year from increasing of car accident. On the other hand, the blood supply can fulfill demand during early December since there is large amount of donor in King's birthday which is 5 December of each year.

Article History: Received 25 March 2018 ◆ Revised 2 June 2018 ◆ Accepted 16 June 2018

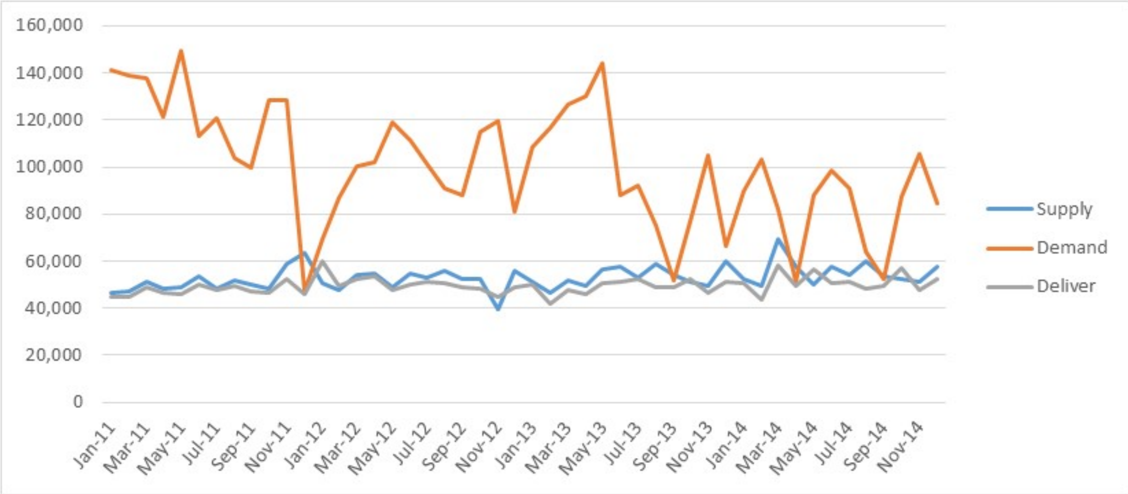

Figure 1. Amount of blood demand, supply, and deliver from Jan. 2011 to Dec. 2014

Blood demand predictions are done by Regions and are often based on experience using heuristic rules that inevitably include a number of biases. Therefore, statistically prediction would be a methodology that can give better and more accurate model for prediction.

Time-series analysis is the process that uses statistical data and measurement to represent characteristics of well-defined data. It is a mature discipline that is used in many research areas such as financial, economics, and engineering. Time-series forecasting is the process of using a model to generate predictions for future events based on known past events. Despite the fact that blood collection and transfusion data is in weekly or monthly basis, forecasting these data using time-series analysis is suitable because it is also capable of predicting seasonal and trend as well.

Nonetheless, blood demands are non-linear time-series data which requires expertise and time-consuming process to predict. For a regular user, automated processes such as raw data import and appropriate prediction model choice together with user friendly visualization tool are necessary for efficient and ease of use system environment.

LITERATURE REVIEW

Pereira et al. (2009) demonstrates that exponential smoothing is more suitable approach than classic ARIMA and neural network. This paper adds two factors to investigate the performance of ARIMAX which is the extension of ARIMA. Two time-series methods/techniques are compared to investigate fitness of the models constructed to predict RBC demands for the NBC and its seven different geolocation Regions. These methods are Holt-Winter's Exponential Smoothing and ARIMAX. The best-fit model will be chosen by comparing the smallest value of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Since blood usage in Thailand is related to amount of Dengue Fever patients who live in diverse humidity geolocations (Nopparat, 2007), this factor will be included into the prediction model as well.

In term of time-series tool development, Cohen recommends R language as suitable technology since it includes many functions for time-series data (Cohen & Cohen, 2008). However, R language alone still lacks of attractive visualization. Additional business intelligence software should be integrated with R in order to provide better visual performance (Lumley, 2006). Tableau is visually oriented and flexible business intelligent software that is also capable of integrating R language into the software with R Script (Underwood, 2013).

PROBLEM DEFINITION/FORMULATION

Since there is no statistically blood demand prediction methodology and tool being used in Thailand, the present blood demand estimation is based on human experience alone. This in turn may lead to inaccurate blood stock and undesired blood shortage. In addition, users may have difficult time to manage uncategorized and large amount of historical data. Therefore, a simple visualization with automated prediction engine will provide great benefit for Thailand blood centers to foresee blood demand and be able to manage activities with relevant data.

Two different time-series prediction methods will be compared to demonstrate the relevant variables of blood demand. Holt-Winter's is the methods without capability of adding independent variables whereas ARIMAX is the model with exogenous variables integration.

Holt-Winter's Exponential Smoothing

Holt-Winter's exponential smoothing enhances the basic exponential smoothing technique with additional capability to support data with trend and seasonal .The technique applies three exponential smoothing formulae to the series. Firstly, the level (or mean) is smoothed to give a local average value for the series; secondly, the trend is smoothed; and lastly, each seasonal sub-series is smoothed .The exponential smoothing formulae applying to a series with trend and constant seasonal component using the Holt-Winters additive technique are as follows:

The forecast for time-period $T + \tau$:

$$\hat{y}_{T+\tau} = \mathrm{a}_T + \tau \mathrm{b}_T + \mathrm{s}_T \tag{1}$$

a_T is the smoothed estimate of the level at time T

b_T is the smoothed estimate of the change in the trend value at time T

s_T is the smoothed estimate of the appropriate seasonal component at T

Autoregressive Integrated with Moving Average and Exogenous Variables (ARIMAX)

The multivariate ARIMAX model expresses an investigated series by incorporating a univariate SARIMA model and related exogenous input series {X}. The relationships between the investigated series and the exogenous series are described by transfer functions. The standard ARIMAX is denoted by ARIMAX(r,n,b)(p,d,q)(P,D,Q)s and is expressed as follows:

$$Q_t^Z = \sum_{j=1}^{U} \frac{\omega_n^j(B)}{\delta_r^j(B)} B^{jb} Q_t^{Xj} + \frac{C + \Theta_Q(B^s)\theta_q(B)e_t}{\Phi_p(B^s)\phi_p(B)\nabla_s^D \nabla^d} \tag{2}$$

Q_t^{Xj} is a blood demand of the jth exogenous input series at time t;

$\frac{\omega_n^j(B)}{\delta_r^j(B)} B^{jb}$ is transfer function of the jth input series;

jb is effect delay of the jth input series;

$\omega_n^j(B) = \omega_0^j - \omega_1^j B^1 - \cdots - \omega_n^j B^n$ is numerator transfer function polynomial of order n

$\delta_r^j(B) = 1 - \delta_1^j B^1 - \cdots - \delta_r^j B^r$ is denominator transfer function polynomial of order r;

Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) is the average absolute error between forecasting and observation data divided by respective observations. It expresses the error as percentage. The measurement is expressed as followed:

$$Q_t^Z = \sum_{j=1}^{U} \frac{\omega_n^j(B)}{\delta_r^j(B)} B^{jb} Q_t^{Xj} + \frac{C + \Theta_Q(B^s)\theta_q(B)e_t}{\Phi_p(B^s)\phi_p(B)\nabla_s^D \nabla^d} \tag{3}$$

Z_t is a blood demand observation at time t;

$\hat{Z}_t$ is a blood demand forecast at time t;

n is a number of observations;

RESEARCH METHODOLOGY

In-depth interview with blood distribution decision maker from National Blood Center is conducted, as well as secondary data of blood distribution from National Blood Center to Regional Blood Centers is collected from NBC database. The secondary data from NBC database is reconciled with NBC yearly report to ensure

```
SCRIPT_REAL("library(forecast);
l<-length(.arg1);
u<-.arg1[1:(l-.arg2[1])];
n<-length(u);u[n]=.arg1[1];
timesr <- ts(u,deltat=1/4,start=c(2012,1));
 fit <- HoltWinters(timesr, beta=FALSE, gamma=FALSE);
fcast <- forecast.HoltWinters(fit, h=.arg2[1]);
append(u,fcast$mean, after = n)",
SUM([NBC Delivered to region]),[PeriodToForecast])
```

Figure 2. R Script in Tableau showing automated Exponential Smoothing

the quality of data. Four years of monthly red blood cell (RBC) demand data from January 2012 to December 2015 are separated into two datasets. 45 points are used to conduct model training that will predict the remaining 3 data points.

Model Fitness Comparison

Two time-series methods are compared to explore appropriate model fitness to predict RBC demand for the National Blood Center (NBC), as well as seven regional blood centers which includes major provinces in Thailand: Region2 covers Central of Thailand, Region3 Eastern of Thailand, Region5 and 6 Northeastern of Thailand, Region8 Central north of Thailand, Region9 Northern of Thailand, and Region12 Southern of Thailand. Holt-Winter's Exponential Smoothing refers to Equation (1) and ARIMAX with Seasonal model refers to Equation (2). Noted that the ARIMAX with seasonality technique is used to construct three sub-models by integrating with two factors: 1) integrating only with the individual platelet demand series (PLT) factor, 2) integrating only with the individual Dengue Fever (DF) infected patients factor, and 3) integrating with both factors, platelet demand series and Dengue Fever infected patient series. The model fitness comparisons are conducted in two sub-steps .Firstly, three ARIMAX sub-models including with ARIMAX with PLT, ARIMAX with DF, and ARIMAX with both PLT and DF are all compared to select the best fit model based on minimum MAPE. Secondly, ARIMAX with lowest MAPE are compared with Exponential Smoothing . Again, the model fitness will be chosen by comparing the smallest value of Mean Absolute Percentage Error (MAPE) refer to Equation (3).

Prediction Tools Development

R Script in Tableau is developed for automated prediction by exponential smoothing method in order to reduce time-consuming for prediction, as well as effective visualization .The script as depicted in **Figure 2** contains exponential smoothing equation which will deliver dynamic result based on input data. The main argument is as followed:

Line 5: Sorting the data; *Line 6*: Fitting the time series (timesr: An object of class, beta: beta parameter of Holt-Winters Filter. gamma: gamma parameter used for the seasonal component); *Line 7*: Forecasting the time series; *Line 8*: Insert the data vectors;

User Acceptance Test

Technology Acceptance Model (TAM) (Davis, 1989) is used for acceptance of the application in term of application benefit, ease-of-use, and need of the application. The authors provide survey to five groups of stakeholders including Head of NBC, head of blood distribution function, head of IT department, head of public relation and donor recruitment, and representative from regional blood centers.

RESULT AND DISCUSSION

The result from model trainings over 45-month time horizon is represented in **Table 1**. ARIMAX with Dengue Fever factor alone performance in RBC demand of NBC represent the best performance comparing with the other two ARIMAX models with lower MAPE. While ARIMAX model with platelet factor alone provides better performance in Region2 and Region5. These two regions are likely to represent the area with scarcity of platelet donors. The RBC predictions of the other 5 regions are best fitted with ARIMAX model with both factors--platelet demand and Dengue Fever patients. The fitness of ARIMAX with both factors is

Table 1. MAPE result from forecasting the RBC demand of NBC and 7 regions over 45-months time horizon by comparing each variable in ARIMAX

Region	ARIMAX(PLT)	ARIMAX(DF)	ARIMAX(PLT)(DF)
NBC	32.35	30.13	33.28
2	27.32	28.58	28.28
3	35.67	49.36	31.55
5	24.60	34.25	33.87
6	27.43	26.51	25.65
8	32.84	54.21	24.09
9	32.63	32.44	31.45
12	28.58	30.21	28.03

Table 2. MAPE result from forecasting the RBC demand of NBC and 7 regions over 45-months time horizon by comparing ARIMAX and Exponential Smoothing

Model	NBC	Region 2	Region 3	Region 5	Region 6	Region 8	Region 9	Region 12
ARIMAX	30.13	27.32	31.55	24.60	25.65	24.09	31.45	28.03
Exponential Smoothing	20.53	21.96	32.40	32.88	23.69	21.77	31.68	24.80

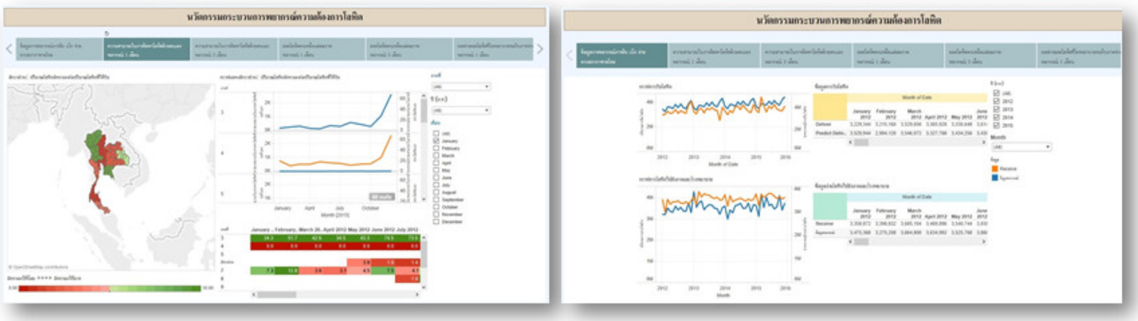

Figure 3. Blood demand prediction visualization in Tableau

reasonably explained by the fact that these areas had severe Dengue Fever epidemics, thus, inevitably required more platelet for treatments.

The result from comparing Exponential smoothing and ARIMAX (**Table 2**) shows that Exponential Smoothing performance at NBC is obviously superior to ARIMAX with lower MAPE. Exponential smoothing performance at Region 2, 6, 8, and 12 also provides better performance fit than ARIMAX. The result indicates that PL and DF may not have high impact on such locations. Nevertheless, ARIMAX provides the best model fit in region 3 and 5 showing a positive sign that PL and DF may have high relevancy to RBC demand in those regions which is reasonable since region 5 and 6 are in the northeast and known to be Dengue Fever epidemic regions of Thailand. The forecasting result for Region 9 shows very close performance between the two methods. Nevertheless, ARIMAX has slightly less error with lower MAPE.

As shown in **Figure 3**, all results are displayed parsimoniously in Tableau, for example, geographical map, line chart, bar chart or pie chart. In addition, multiple visuals can be combined into one screen known as a dashboard.

Figure 4 display the Need of Application, Ease-of-use, and Benefit of Application from five stakeholder groups. As a result, the application is highly needed along with easy to use with around 93% positive feedback. However, the application still have some gap to develop in order to give full benefit to users.

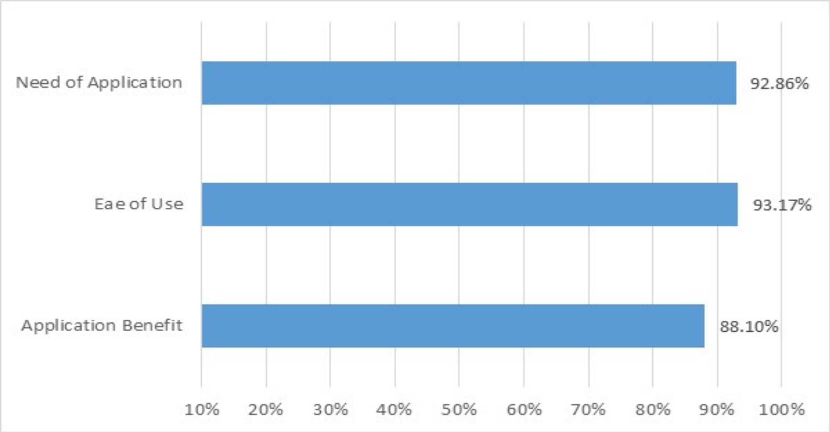

Figure 4. Result from Technology Acceptance Survey

CONCLUSION

The research provides statistical blood demand prediction as a guideline to foresee blood need in order to prepare blood supply based on different geolocations. Different blood-bank locations require different prediction models .The diverse techniques found in RBC demand model fitting are likely to cause by different climate or epidemic in each location. Also, population and sociological profiles in each location must be taken into consideration. For example, some locations may have people who are willing to donate blood whereas other locations are lack of donors. The shortage of donors may cause that location to request a large amount of blood from the NBC. The resulting models can be used to develop a guideline for blood transfusion prediction because they have incorporated both seasonality and relevant factors along with simple visualization and automated prediction process. Since blood is one kind of perishable products, systematic and automated blood demand prediction tool does benefit not only the NBC and other regional blood banks but also the hospitals to be able to plan and manage blood transfusion with ease of use. However, other relevant factors such as accident rate, deceases, and logistic path should also be considered as they may have impacts on blood demand .Future research should also include relationships among all parties involved in blood transfusions, i.e. donors, logistics, distribution centers, and hospitals. In addition, the future blood prediction model and application should also include separation of regular blood type and rare blood type.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Ardthawee Tanyavutti – Technopreneurship and Innovation Management Program (Chulalongkorn University), Thailand.

Uthai Tanlamai – Department of Accountancy (Chulalongkorn University), Thailand.

REFERENCES

Cohen Y., & Cohen J. Y. (2008). Statistics and Data with R: An applied approach through examples, 251-277

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*(3), 319. https://doi.org/10.2307/249008

Lumley T. (2006). R Fundamentals and Programming Techniques. *R manual*, 1-225.

Nopparat, N., & Nopparat, S. (2007). Seasonal climatic variation and dengue hemorrhagic fever in Uttaradit Province. HSRI Journal*, 1*(1), 68-79.

Owens, W., Tokessy, M., Rock G. (2001). Age of blood in inventory at a large territory care hospital. *Vox Sanguinis, 81*, 21-23. https://doi.org/10.1046/j.1423-0410.2001.00054.x

Pereira, A. (2004). Performance of time-series methods in forecasting the demand for red cell Transfusion. *Transfusion Practice, 44*, 739-746. https://doi.org/10.1111/j.1537-2995.2004.03363.x

Perera, G., Hyam C., Taylor C., & Chapman J. F. (2009). Hospital blood inventory practice: the factors affecting stock level and wastage. *Transfusion medicine, 19*, 99-104. https://doi.org/10.1111/j.1365-3148.2009.00914.x

Underwood, J. (2013). *Predictive Data Visualization with Tableau, BI and Advance Analytics* Available at: http://www.jenunderwood.com/category/tableau/