BLOOD DONORS & BLOOD COLLECTION

TRANSFUSION

A robust autonomous method for blood demand forecasting

Esa V. Turkulainen[1] [iD] | Merel L. Wemelsfelder[2] [iD] | Mart P. Janssen[2] [iD] |
Mikko Arvas[1] [iD]

[1]Research and Development, Finnish Red
Cross Blood Service, Helsinki, Finland

[2]Transfusion Technology Assessment
Group, Donor Medicine Department,
Sanquin Research, Amsterdam, the
Netherlands

Correspondence
Esa V. Turkulainen, Finnish Red Cross
Blood Service, Helsinki, Finland.
Email: esa.turkulainen@veripalvelu.fi

Abstract

Background: Blood supply chain management requires estimates about the demand of blood products. The more accurate these estimates are, the less wastage and fewer shortages occur. While the current literature demonstrates tangible benefits from statistical forecasting approaches, it highlights issues that discourage their use in blood supply chain optimization: there is no single approach that works everywhere, and there are no guarantees that any favorable method performance continues into the future.

Study Design and Methods: We design a novel autonomous forecasting system to solve the aforementioned issues. We show how possible changes in blood demand could affect prediction performance using partly synthetic demand data. We use these data then to investigate the performances of different method selection heuristics. Finally, the performances of the heuristics and single method approaches were compared using historical demand data from Finland and the Netherlands. The development code is publicly accessible.

Results: We find that a shift in the demand signal behavior from stochastic to seasonal would affect the relative performances of the methods. Our autonomous system outperforms all examined individual methods when forecasting the synthetic demand series, exhibiting meaningful robustness. When forecasting with real data, the most accurate methods in Finland and in the Netherlands are the autonomous system and the method average, respectively.

Discussion: Optimal use of method selection heuristics, as with our autonomous system, may overcome the need to constantly supervise forecasts in anticipation of changes in demand while being sufficiently accurate in the absence of such changes.

Abbreviations: APE, absolute percentage error; ARIMA, autoregressive integrated moving average; AUTO-N, Autonomous (method Selection) forecasting method with N-step selection period length; AVG, method average; COVID-19, coronavirus disease 2019 caused by SARS-CoV-2; DYNREG, dynamic regression; ELM, extreme learning machines; ETS, exponential smoothing method; MA, moving average; MAPE, mean absolute percentage error; MLP, multilayer perceptrons; NNAR, autoregressive neural networks; SNAIVE, Seasonal Naïve Method; STL, season-trend decomposition method; STLF, multiseason-trend decomposition method; TBATS, exponential smoothing method with trigonometric seasonality, box-cox transformation, ARIMA errors, trend, and seasonal components; W.AVG, weighted method average.

1 | INTRODUCTION

A well-functioning blood supply chain operation reduces wastage, procurement costs, and avoids life-threatening shortages. Research on blood supply chain efficiency and reliability has increased steadily over the past couple of decades. According to a recent survey,[1] most research on supply chain planning assumes demand as stochastic, deterministic, or otherwise intractably uncertain, using constant estimates or distribution sampling. However, demand forecasting has been deemed necessary and superior to expert planning in several studies,[1–11] suggesting that supply chain management approaches can be improved by a significant margin by adopting methods for demand forecasting. Almost all reviewed research on demand forecasting attempts to determine the single best method for reducing shortages or costs, revealing that the best method varies between blood banks and blood products.[3,12–14] Researchers also note on multiple occasions that there are no particular methods that are guaranteed to give accurate forecasts and that the best method may change over time.[3,12,14] To avoid laborious periodic method reselection, others have suggested using method selection systems and specific automatic procedures (most often the Box-Jenkins procedure[15]) to help evaluate and select Autoregressive Integrated Moving Average (ARIMA) models without human intervention.[4–6,13,14]

In this study, we use a marked but relatively brief change in the behavior of the weekly red blood cell demand in Finland to examine how the continuation of this behavior would affect method performance and selection. We then devise heuristics for method selection that could allow an autonomous system to produce robust forecasts without supervision during changes in the demand behavior and compare the performance of such a system with individual method performance using altered demand data. Finally, we compare the performances of both the heuristics and the individual methods using unaltered demand data from Finland and the Netherlands to gauge its real-world applicability.

2 | MATERIALS AND METHODS

2.1 | Data

Both the Finnish Red Cross Blood Service and Sanquin record every unit of blood delivered. By filtering these data by specific blood products (e.g., red blood cells or platelets) and aggregating them per week, we can create a weekly demand series. The record of red blood cell product deliveries spans from 2014 to 2021 in Finland and from 2009 to 2020 in the Netherlands. The aggregated weekly (Monday to Sunday) demand from this period for both countries is shown in Figure 1. The Finnish series exhibits a seemingly significant change in behavior around 2018 and 2019, as the slight downtrend in demand seems to level out and create a more pulse-like signal. This behavior, however, does not continue beyond 2019. To study how the performance of different methods might be affected by such changes, we artificially extend the anomalous behavior by modeling the series between July 2017 and July 2019, adding some noise to it, and overwriting the actual demand history from July 2019 onward. The process to generate the synthetic data is explained in detail in Appendix S1, Part A. The artificially extended demand history is shown in Figure 2. We use both the altered and unaltered version of the demand data to test forecasting methods.

2.2 | Methods

As most blood supply operators do not have access to meaningful clinical variables or other external regressors, we limit ourselves in this study to autoregressive forecasting methods to enable a broad applicability of the results. Based on the current literature, we chose to examine the following methods: simple moving averages (MA), exponential smoothing (ETS), ARIMA, and autoregressive neural networks (NNAR). In addition to these, we examined the seasonal naïve method (SNAIVE), method averaging (AVG), season-trend decomposition methods (STL and STLF), a dynamic seasonal method (TBATS), dynamic regression (DYNREG), multilayer perceptrons (MLP), and extreme learning machines (ELM). All methods are explained in detail in Appendix S1, Part B.

The methods are trained using 3 years of most recent demand history. A forecast is generated for the following week, and its accuracy is tested using absolute percentage errors (APEs). This process is repeated until we have backtested all of the available history. These tests allow us then to summarize method performances in mean absolute percentage errors (MAPEs):
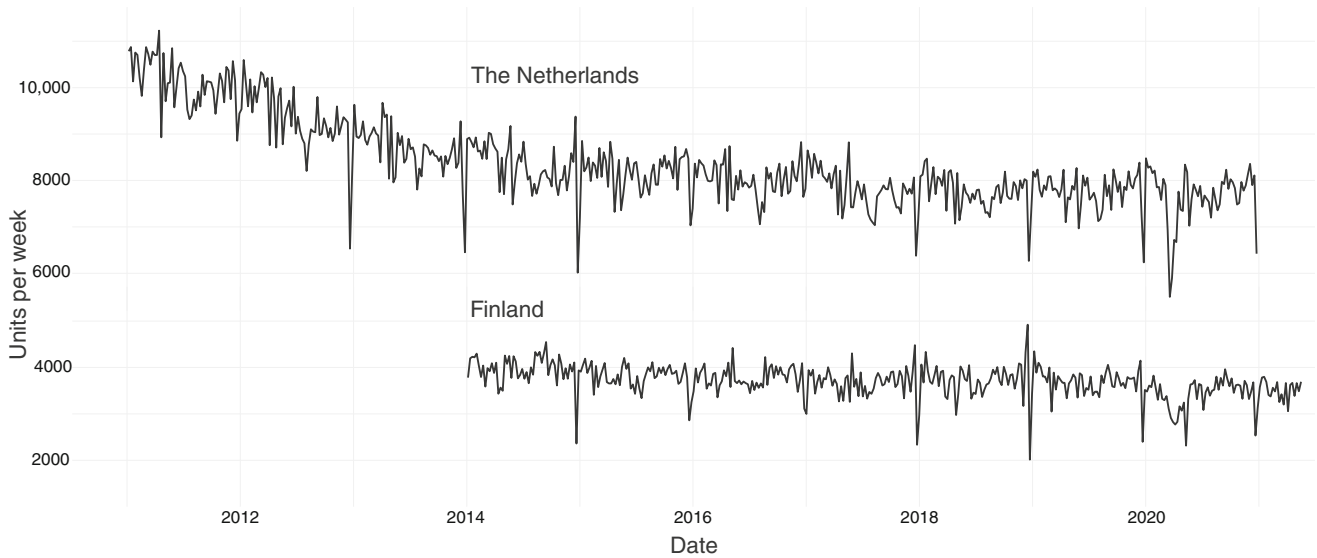
FIGURE 1 Weekly demand of red blood cell products in the Netherlands and Finland
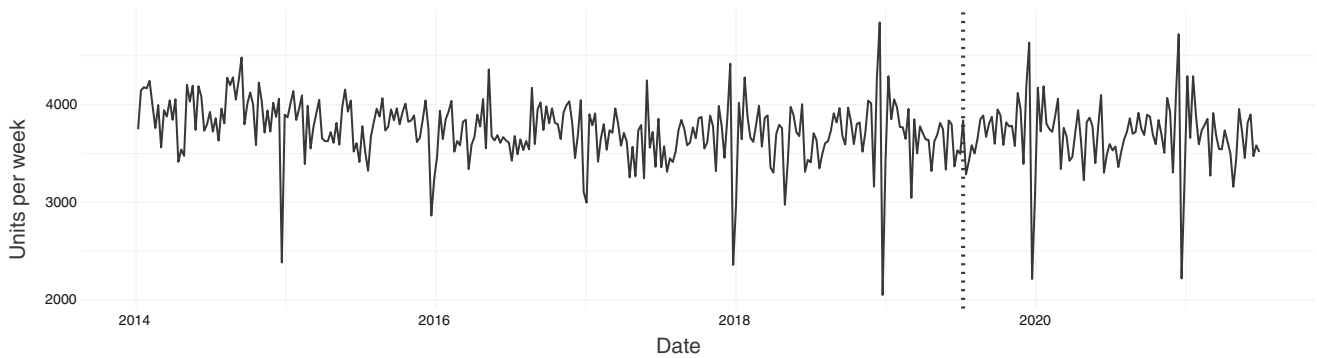

FIGURE 2 Altered weekly demand of red blood cell products in Finland. The dotted line indicates the beginning of the synthetic data

$$\text{MAPE} = 100 * \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \widehat{y}_i}{y_i} \right|,$$

where $\widehat{y}_i$ denotes the forecast value and y_i denotes the observed value for a specific week i.

Absolute percentage errors are unit-free, so they enable comparisons between different data sets and scales. However, they have some important limitations. Division by zero issues arise when the target observations contain zeroes, and the metric becomes very unstable if data has values very close to zero. Additionally, the data should exist on a ratio scale (entirely positive, meaningful zero). Our weekly aggregates of blood demand do not contain zeroes and the data are positive. We discuss other possible method selection metrics in the Discussion section. All development code is written in R (version 4.0.5).[16] All methods are implemented using the *forecast*

package (version 8.13)[17] and the *nnfor* package (version 0.9.6),[18] except for the combination methods (AVG and W.AVG). The development code is publicly available on Github.[19]

The repository contains all scripts used for generating the results in this article, an implementation of an autonomous forecasting system (an R Markdown script that outputs an HTML forecast report), as well as the necessary custom functions for running such a system. No data are made available for confidentiality reasons.

3 | AUTONOMOUS METHOD SELECTION SYSTEM

Most of the methods, as implemented in the *forecast* package, are so-called "modelers," meaning that a new

model is fit every time a forecast is requested. This allows for a degree of adaptability regarding changes in series behavior. However, all modelers are restricted to their respective model formulations, limiting more general adaptability. This issue can be circumvented by selecting from a pool of different modelers, thereby expanding the model space available for selection. As periodic manual reselection is costly, and often unfeasible, autonomous application of predetermined selection heuristics may be preferable. Figure 3 illustrates one possible method selection process for an autonomous forecasting system, an implementation of which is provided on Github.[19] To test whether allowing for autonomous method selection using historical accuracy metrics would yield advantages

in accuracy over the individual methods and modelers, we selected forecasts from the methods that held the lowest MAPE in 1, 3, 6, 12, 18, and 24 weeks before the forecast date (the selection period). We also tested forecasting by ranking methods based on the previous week's performance and computing an exponentially decaying weighted average as the forecast (W.AVG).

4 | RESULTS

Figure 4 shows the differences in prediction accuracy between the first and second half of *the artificially extended weekly data* (Figure 2). We see that the STL, STLF, and
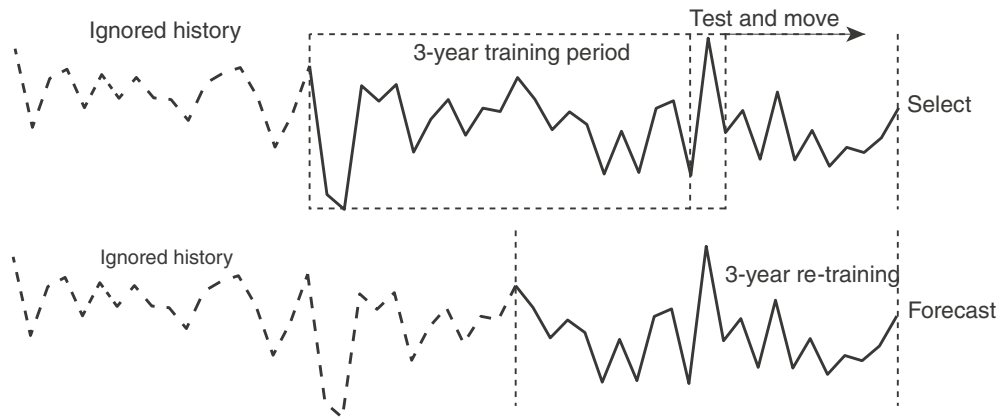
FIGURE 3 A schematic of the method selection process. Methods are trained with a 3-year training window and then tested on the subsequent observation. The training window then moves, dropping the oldest observation, and including the previous test observation. A series of tests are performed over a selection period up until the forecast date (e.g., 12 weeks). Next, a forecast is generated by selecting the method with the lowest mean absolute percentage error (MAPE) from the selection period and retraining it with 2 years of the most recent observations
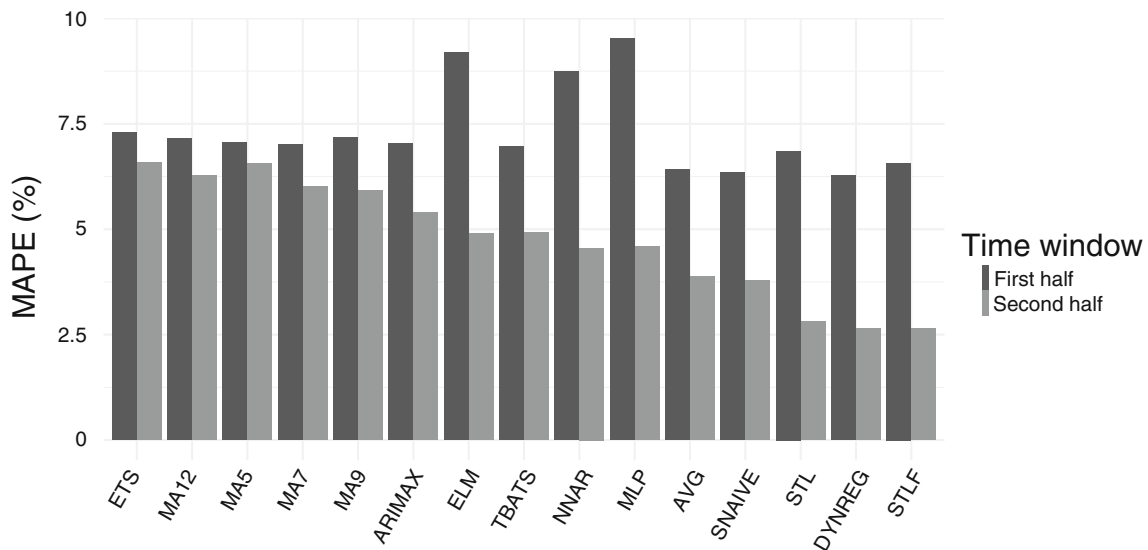
FIGURE 4 Method performances on the first and second half of the synthetic data, ordered by the error magnitude on the second half. Some methods benefit from the change, some suffer

DYNREG methods gain a significant advantage over other methods in the latter half of the data, after first performing largely comparably to other methods. The neural network methods (NNAR, MLP, ELM) stand out here in their substantial increase in accuracy in the latter half of the data after being the clear underperformers in the first half. However, they still compare unfavorably with AVG, SNAIVE, STL, DYNREG, and STLF. Table 1 shows the overall mean absolute percentage errors of different methods, and Table 2 shows the overall mean absolute percentage errors of method selection heuristics for different testing periods as well as the result for the weighted averaging method. The best accuracies in both Tables are highlighted in bold. Among the individual methods, the DYNREG performs best (4.46%). The majority of selection heuristics examined outperform most individual methods, with AUTO-12 at the top (4.33%), beating them all. *For the unaltered demand data* (Figure 1), the performances of single methods and selection heuristics are presented together as a heatmap in Figure 5. The first row presents the theoretical minimum selection error, found by selecting the most accurate forecast each week retrospectively. The results are separate for each length of the testing period, as extending this period limits the available data, thus reducing the number of errors that can be computed.

For Finland, the best performers are the AUTO-12 selection method and AVG. The selection methods compared favorably with most individual methods, indicating viability for practical use without loss in accuracy. In the Netherlands, the accuracies are much better, most likely due to the larger volume of demand decreasing the noisiness of the signal. The best method was AVG, with STLF, ARIMAX, and DYNREG performing almost equally well. Out of the selection methods, the W.AVG and AUTO-6 performed best, with better or roughly equal accuracy when compared with the individual methods. Overall, differences between accuracies were not statistically significant at the 95% confidence level, except for the MLP, whose underperformance was significant in comparison with some methods even after correcting for multiple testing.

5 | DISCUSSION

The aim of this article was to explore blood demand forecasting systems that could offer robustness in the event of changes in the underlying demand signal. For this purpose, we identified a period of unusual behavior in the Finnish weekly blood demand data and extended it to create a synthetic demand history. We then examined how these conditions would affect the performance of various popular forecasting methods and found that the

TABLE 1 Overall mean absolute percentage errors (MAPEs) of methods over the entire synthetic data

SNAIVE	MA-5	MA-7	MA-9	MA-12	ETS	STL	TBATS	ANN	ARIMAX	DYNREG	STLF	MLP	ELM	AVG
5.06	6.81	6.51	6.54	6.72	6.94	4.84	5.94	6.64	6.21	**4.46**	4.61	7.06	7.04	5.14

Note: The best accuracies in both Tables are highlighted in bold.

AUTO-1	AUTO-3	AUTO-6	AUTO-12	AUTO-18	AUTO-24	W.AVG
5.01	4.64	4.54	**4.33**	4.68	4.71	4.76

TABLE 2 Overall mean absolute percentage errors (MAPEs) of selection heuristics over the entire synthetic data

Note: The best accuracies in both Tables are highlighted in bold.

	Finland						The Netherlands						
	1	3	6	12	18	24	1	3	6	12	18	24	
	2.08	2.02	2.03	2.06	2.05	1.95	0.99	0.99	0.99	0.99	0.98	0.99	MIN
	7.50	7.49	7.50	7.57	7.54	7.49	6.16	6.15	6.15	6.16	6.07	6.07	SNAIVE
	6.54	6.50	6.51	6.58	6.63	6.58	4.00	4.00	4.00	4.03	4.02	4.02	MA-5
	6.36	6.28	6.29	6.35	6.40	6.36	4.06	4.05	4.06	4.08	4.07	4.08	MA-7
	6.52	6.44	6.45	6.52	6.57	6.54	4.07	4.07	4.08	4.09	4.07	4.08	MA-9
	6.77	6.70	6.71	6.81	6.87	6.85	4.22	4.22	4.23	4.25	4.23	4.24	MA-12
	6.94	6.87	6.89	7.01	7.08	7.05	4.17	4.17	4.18	4.20	4.17	4.18	ETS
	6.78	6.74	6.75	6.75	6.81	6.74	4.62	4.61	4.61	4.64	4.63	4.63	STL
	6.68	6.64	6.66	6.67	6.70	6.64	3.92	3.91	3.91	3.93	3.93	3.93	STLF
	7.16	7.08	7.11	7.21	7.27	7.25	4.02	4.01	4.01	4.03	4.02	4.04	ARIMAX
	6.62	6.59	6.60	6.60	6.66	6.63	3.96	3.94	3.95	3.97	3.97	3.98	DYNREG
	7.01	6.96	6.98	7.11	7.19	7.16	4.22	4.22	4.24	4.26	4.24	4.26	TBATS
	7.57	7.53	7.56	7.62	7.63	7.60	4.79	4.78	4.79	4.82	4.79	4.79	NNAR
	9.50	9.47	9.53	9.65	9.70	9.68	6.23	6.18	6.18	6.22	6.18	6.21	MLP
	8.37	8.33	8.35	8.44	8.51	8.35	5.51	5.47	5.46	5.48	5.43	5.43	ELM
	5.99	5.93	5.94	5.99	6.03	5.96	3.74	3.73	3.74	3.76	3.74	3.74	AVG
	6.21	6.25	6.26	6.30	6.31	6.28	4.00	4.01	4.01	4.03	4.01	3.99	W.AVG
	6.31	6.36	6.35	6.42	6.42	6.39	4.36	4.36	4.37	4.38	4.36	4.35	AUTO-1
	NA	6.25	6.30	6.42	6.44	6.41	NA	4.23	4.23	4.23	4.22	4.21	AUTO-3
	NA	NA	6.23	6.28	6.21	6.28	NA	NA	4.08	4.06	4.06	4.08	AUTO-6
	NA	NA	NA	5.85	5.88	6.00	NA	NA	NA	4.18	4.19	4.20	AUTO-12
	NA	NA	NA	NA	6.55	6.63	NA	NA	NA	NA	4.23	4.25	AUTO-18
	NA	NA	NA	NA	NA	6.70	NA	NA	NA	NA	NA	4.13	AUTO-24

FIGURE 5 A heatmap presentation of the overall mean absolute percentage errors (MAPEs) for the methods and selection heuristics. The first row presents the error for the optimal selection strategy (theoretical minimum). Results are spread over six columns (for each testing period length) to ensure comparable error scores. The splits separate countries and individual methods from heuristics

method performance is greatly affected by structural changes in the demand, highlighting the need to reselect methods over time (Figure 4). To test whether this could be done autonomously, we devised two different method selection heuristics that rely on historical performance: one that chooses the method that had the best mean absolute percentage error over a certain predetermined length of history and one that creates an exponentially weighted average of forecasts from different methods. The results from an application of various methods for the synthetic data show dominance of the selection heuristics over all individual methods, suggesting that the ability to switch between methods autonomously or to decide their weighting within a combination of methods successfully utilizes the differences between the methods. The method selection process behind these results is illustrated in Figure S1. While our study does not exhaust the search for the optimal selection strategy, the fact that a large majority of the heuristics outperform all of the individual methods indicates that there are multiple ways to achieve superiority in accuracy over single method approaches.

Finally, we tested the heuristics and individual methods using the real, unaltered weekly demand data from Finland and the Netherlands. The results indicate that the heuristics' advantage with the altered demand series is most likely a result of the persistence of the behavioral change, as in reality, they do not differ significantly in accuracy from the individual methods in either countries. The AVG performed strongly in both countries, which is likely due to a trade-off between the robustness enabled by the constant method reselection and simply forecasting with the historically best individual method. The AVG method, which computes the mean of all individual forecasts, seems to take advantage of the random error in the predictions of the individual methods. By canceling the biases of the individual methods, it reduces the variance of the resulting forecast. However, when more structural changes in the underlying demand signal emerge, the averaging method is unable to leverage the advantages gained by some methods or minimize the penalties suffered by others, as is shown by the dominance of selection heuristics for the synthetic data.

While it is impossible to build an autoregressive forecasting system that can foresee changes in demand behavior (for example, a global COVID-19 pandemic essentially halting regularly scheduled surgeries), one can leverage the advantages of various individual methods by enabling selection from a diverse library of methods. We did not aim to find the optimal set of methods nor the absolute best or most generalizable rule set for selection. For example, in an operation critical environment such as the blood supply chain, it might be more meaningful to select methods using a metric that harshly penalizes underforecasting or by their ability to detect peaks. We found no statistically significant differences between the peak detection abilities of different methods (Appendix S1, Part D). Nevertheless, by utilizing some simple method selection heuristics, one can ensure that any changes in demand behavior will be adjusted for, while maintaining sufficient forecasting accuracy in the absence of such changes. As such, we conclude that this approach is a viable autonomous forecasting solution for blood centers and blood products in general.

CONFLICT OF INTEREST
The authors have disclosed no conflicts of interest.

ORCID
Esa V. Turkulainen https://orcid.org/0000-0001-8844-9344
Merel L. Wemelsfelder https://orcid.org/0000-0001-5421-3392
Mart P. Janssen https://orcid.org/0000-0002-1682-7817
Mikko Arvas https://orcid.org/0000-0002-6902-8488

REFERENCES
1. Pirabán A, Guerrero WJ, Labadie N. Survey on blood supply chain management: models and methods. Comput Oper Res. 2019;112:104756.
2. Frankfurter GM, Kendall KE, Pegels CC. Management control of blood through a short-term supply-demand forecast system. Manage Sci. 1974;21(4):444–52.
3. Pereira A. Performance of time-series methods in forecasting the demand for red blood cell transfusion. Transfusion. 2004;44(5):739–46.
4. Silva Filho OS, Cezarino W, Salviano GR. A decision-making tool for demand forecasting of blood components. IFAC Proc Vol. 2012;45(6):1499–504.
5. Silva Filho OS, Carvalho MA, Cezarino W, Silva R, Salviano G. Demand forecasting for blood components distribution of a blood supply chain. IFAC Proc Vol. 2013;46(24):565–71.
6. Fortsch SM, Perera S. A resilient donor arrival policy for blood. Oper Res Health Care. 2018;19:165–74.
7. Kumari DMS, Wijayanayake AN. An efficient inventory model to reduce the wastage of blood in the national blood transfusion service. 2016 Manuf Ind Eng Symp; 2016 Oct 22; Colombo, Sri Lanka. IEEE; 2016 Dec 26.
8. Guan L, Tian X, Gombar S, Zemek AJ, Krishnan G, Scott R, et al. Big data modeling to predict platelet usage and minimize wastage in a tertiary care system. Proc Natl Acad Sci USA. 2017;114(43):11368–73.
9. Khaldi R, El Afia A, Chiheb R, Faizi R. Artificial neural network based approach for blood demand forecasting: fez

transfusion blood center case study. Proceed 2nd Intl Conf Big Data Cloud Appl; 2017 Mar 29-30; Tetouan, Morocco. ICPS Proceed; 2017;59:1–6.

10. Ahmadimanesh M, Tavakoli A, Pooya A, Dehghanian F. Designing an optimal inventory management model for the blood supply chain: synthesis of reusable simulation and neural network. Medicine. 2020;99(29):e21208.

11. Li N, Chiang F, Down DG, Heddle NM. A decision integration strategy for short-term demand forecasting and ordering for red blood cell components. Oper Res Health Care. 2021;29: 100290.

12. Gardner ES. Box-Jenkins vs multiple regression: some adventures in forecasting the demand for blood tests. Interfaces. 1979;9(4):49–54.

13. Fanoodi B, Malmir B, Jahantigh FF. Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. Comput Biol Med. 2019;113:103415.

14. Shih H, Rajendran S. Comparison of time series methods and machine learning algorithms for forecasting Taiwan Blood Services Foundation's blood supply. J Healthc Eng. 2019;2019: 6123745.

15. Box GE, Jenkins GM, Reinsel G. Time series analysis: forecasting and control. San Francisco: Holden-day; 1970.

16. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

17. Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, et al. forecast: forecasting functions for time series and linear models. Version 8.13. 2020. Available from: https://pkg.robjhyndman.com/forecast/

18. Kourentzes N. nnfor: time series forecasting with neural networks. R package version 0.9.6. 2019. Available from: https://CRAN.R-project.org/package=nnfor

19. Finnish Red Cross Blood Service. production_forecasts. Github. 2021. Available from: https://github.com/FRCBS/production_forecasts

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Turkulainen EV, Wemelsfelder ML, Janssen MP, Arvas M. A robust autonomous method for blood demand forecasting. Transfusion. 2022;62:1261–8. https://doi.org/10.1111/trf.16870